

Application of Curve Registration Methods on Analyzing Wearable Device Data

Hainiu (Johnny) Xu

Supervisors: Prof. Jane-Ling Wang

Submitted in partial fulfillment of the requirement for the
Honors degree in B.S. Statistics

Department of Statistics
University of California, Davis
June 12, 2020

Contents

1	Introduction	1
2	Literature Review	5
2.1	Overview	5
2.2	Landmark Approach	6
2.3	Procrustes Approach	8
2.3.1	Dynamic Time Warping and the Procrustes Fitting Criterion	8
2.3.2	Registration Using Fisher-Rao Metric	10
2.4	Pairwise Approach	14
2.5	Other Time Warping Methods	16
3	Data Analysis	18
3.1	Introduction	18
3.2	Landmark Registration	20
3.2.1	1-Step Landmark Registration	20
3.2.2	2-Step Landmark Registration	26
3.3	Pairwise Curve Synchronization	27
3.3.1	1-Step Pairwise Curve Synchronization	27
3.3.2	2-Step Pairwise Curve Synchronization	29
3.4	Warping with Fisher-Rao Metric	30
3.4.1	1-Step Fisher-Rao Warping	30
3.4.2	2-Step Fisher-Rao Warping	32
3.5	Discussion of Data Analysis Result	33
3.6	Conclusion of Data Analysis	37
3.6.1	Landmark Registration and Pairwise Curve Synchronization	38
3.6.2	Fisher-Rao Warping	39

Acknowledgements

First of all, I would like to express my utmost appreciation to Professor Jane-Ling Wang, who has not only being a superior supervisor throughout my researches but also a life mentor. It is my great honor to have Professor Wang assisting me with my thesis writing and my understanding of topics in the field of statistics. Professor Wang has set a great example of a mentor, a researcher, and a learner that I wish to grow up to one day.

I would also like to pay my special regards and gratitude to Cody Carroll, who is currently a Ph.D. candidate at the department of statistics of UC Davis. It is my privilege to have Cody helping me with my understanding of the research topics as well as the thesis writing. Cody's knacky writing skill and transpicuous explanation of thesis-related topics have played an indispensable role in leading me through the journey of thesis writing.

In addition, I wish to express my gratitude to the faculty and staff at the statistics department. The statistics Research Training Group funded by the National Science Foundation and organized by Professor Wolfgang Polonik and other faculties and staff at the department exposed me to research and provided me the invaluable chance to learn research skills through conducting interesting research projects. Specifically, the project, analysis of data collected from wearable devices, designed by Professor Jane-Ling Wang led me into the realm of functional data analysis and edified me to work on this very thesis. Ms. Kimberly McMullen, who is the undergraduate program coordinator, has also been supportive. Kim first introduced the Research Training Group as well as the honors thesis program to me. Without her timely and crucial information, I would have missed such a vital opportunity. Finally, I feel lucky to have been a student of most faculties in the department. Their diverse lectures and enlightening office hours helped me to build and strengthen my foundation in statistics.

Last but not the least, I would like to recognize the invaluable assistance that my family provided during my study. I am fortunate to have my father, Bo Xu, and mother, Jing Zhang, standing behind me and always willing to be my sheltering harbor. Their endless love and support allowed me to fearlessly chase my dream. I would also like to give my wholehearted appreciation to Yuchen Si, who have been an adorable, adventurous, and upbeat girlfriend of mine for four years and many years to come. Yuchen's understanding and support gave me the confidence and strength to dive into the thesis and put my best effort into it.

The future belongs to those who believe in the beauty of their dreams. May we keep the faith in our dreams and prosper in the future.

1 Introduction

With the advancement in technology, electronic devices are becoming more and more compact, which allows numerous devices to be worn on human body. These types of devices are often referred to as *wearable devices*. Until recently, wearable devices have mainly been seen in the medical research field and health care industry, for the purpose of recording and monitoring people's health data. In recent years, wearable devices have found their way to the public consumer market and the market for wearable devices has expanded significantly thereafter [1]. Among the wearable devices, most of them have been made as smart watches or wrist band like Apple Watch, Galaxy Watch, Fitbit, etc. Other kinds of wearable devices take the form of glasses or rings.

Since wearable devices are designed to be worn on the body continuously, these devices are capable of frequently recording time-varying data like the heartbeat, step count, calorie consumption, sleep quality, etc. Such type of device is typically named healthcare wearable device, which is also the mainstream of wearable devices such as Fitbit and other wearable monitors [2]. Healthcare devices have been used widely to conduct numerous researches including using wearable devices to help Parkinson patients, using wearable device accelerometer data to conduct activity classification, etc. [3, 4] In this case, instead of treating wearable device data as discrete with very dense recordings, which is what most previous researchers have done, the rapidly observed recordings may be viewed collectively as observations of continuous functions over time. Thus, the prosperity of the consumer wearable devices has led to new functional data and these data demand a systematic way to be analyzed. Functional data analysis (FDA) is the statistical branch designated to study and analyze functional data. Functional data are usually assumed to be smooth and continuous curves and the statistical analysis of such curves is often referred to as FDA.

FDA have been studied extensively and the major principles and methods for FDA are discussed in numerous monographs including *Functional Data Analysis* by Ramsay and Silverman as well as *Introduction to Functional Analysis* by Kokoszka and Reimherr [5, 6].

Since wearable device data can also be treated as functional data, the FDA methods becomes the felicitous tool to be utilized to analyze data from wearable devices. The specific FDA tools that is extensively utilized in this study are Functional Principal Component Analysis (FPCA) and curve registration. FPCA, similar to the multivariate principal component analysis, reduces the dimension of the infinite dimensional functional data to a finite dimension by decomposing each individual function to its cross-sectional mean and the spectral decomposition of the covariance surface of a sample of functions. Such decomposition is named the Karhunen-Loève decomposition, which is precisely defined as

$$X_i(t) = \mu(t) + \sum_{k=1}^{\infty} A_{ik} \phi_k(t) \quad (1)$$

where $X_i(t)$ is the i^{th} realization of a square-integrable stochastic process X , $\mu(t)$ represents the cross-sectional mean of the sample curves, A_{ik} is the k^{th} functional principal components of X_i , and $\phi_k(t)$ are the k^{th} corresponding eigenfunction of the auto-covariance operator [7, 8]. The auto-covariance operator is defined as

$$\Sigma(g) = \int_{\tau} \Sigma(s, t) g(s) ds$$

where $g \in \mathbb{L}^2$, $s, t \in \tau$ and τ is the domain of the sample functional data and it has spectral decomposition

$$\Sigma(s, t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t)$$

Therefore, functions can be well-represented by the several leading principal components that explain the major mode of variation between functions. Since FPCA is mainly based on the spectral decomposition of the covariance surface, its is therefore highly depended on the integrity of the covariance surface and the cross-sectional mean of the observed functions. Furthermore, notice that the cross-sectional mean requires that the observed data are aligned to make it representative or it will be distorted by the peak or valleys of the misaligned curve,

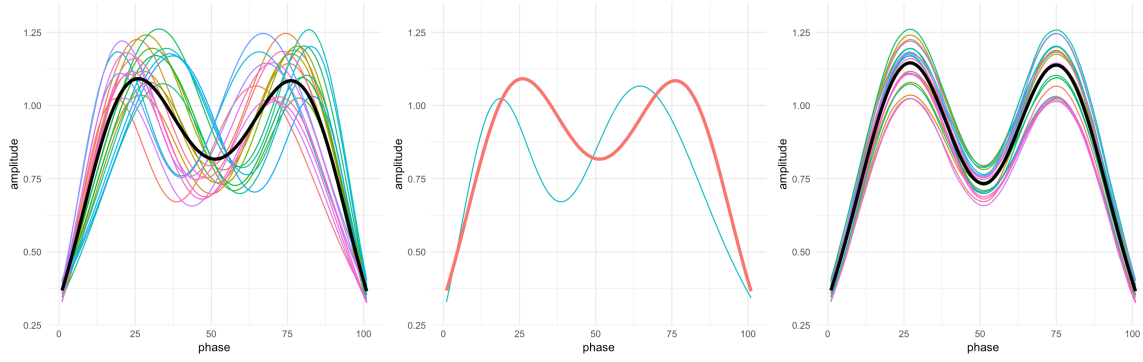


Figure 1: Demonstrate of misaligned data and its effects on the cross-sectional mean. Left: misaligned functional data and their cross-sectional mean (thick black line). Middle: The misaligned cross-sectional mean and a sample curve. See how it is unrepresentative to the population. Right: Aligned curves and their cross-sectional mean (thick black line). Simulation data generated by Srivastava et. al [10]

which is illustrated in Figure 1. In addition, the sample covariance function is the stochastic process X is defined as

$$C(s, t) = \mathbb{E}[(X(s) - \mathbb{E}[X(s)])(X(t) - \mathbb{E}[X(t)])]$$

Hence, the decomposition only capture variation in the vertical direction and it could be easily influenced the other kinds of variations. Therefore, to get a prominent result from FPCA, one must pay close attention to the type of variation presented in the dataset. In the case of wearable device data, most data have two types of variation presented— one of them is the difference in magnitude between functions, which is often referred to as the *amplitude variation*, and the other form of variation is presented in the horizontal time axis of the function, which is called the *phase variation* [9]. Based on the FPCA framework, the presence of phase variation could easily lead to inefficient estimates and thus lead to inefficient eigenvalues and eigenfunctions.

Intuitively, the amplitude variation of wearable device data reflects the difference in the magnitude of the corresponding variable across different users. For instance, the amplitude variation of the calorie consumption variable reflects the difference in the amount of calorie burned between users. Such amplitude variation occurs mainly due to the fact that every

user have her own living style and different activity intensity and therefore the measurement of their activity data would differ in magnitude. Such amplitude variation can be well explained by FPCA. The same interpretation also goes for phase variation. In the context of wearable device data, phase variation occurs also due to differences in users' daily activity. For instance, the calorie data could display phase variation if one user workouts at 2 pm and the other user works out at 4 pm. Under this circumstance, the calorie recording will reach its peak at 2 pm for one user and the other user's calorie recording will peak at 4 pm, causing phase variation. Such The phase variation obscures the amplitude variation by "stretching" or "shrinking" the function in the horizontal direction and thus misalign the amplitude features (peaks, valleys, etc.). For instance, the plot in Figure 2 shows how two functions, albeit each curve has only one peak, the peaks of each function occurred at completely different locations in the horizontal axis due to phase variation, which resulted in the cross-sectional mean been distorted and the mean function almost look like it has two peaks. Therefore, in order to get a prominent result from FPCA, which is only capable of analyzing amplitude variation, one must first separate amplitude and phase variation so that the phase variation can be eliminated from the data. The process of eliminating phase variation is called *curve registration* (or time warping in the field of engineering). Thankfully, methods for curve registration have been studied extensively by multiple research groups and numerous different curve registration methods have been proposed. In this study, the result of different curve registration methods will be analyzed using a Fitbit dataset, which contains step-count data of 35 users. The Fitbit dataset will be introduced in detail in the data analysis section.

The thesis is organized as follows: the literature review of the major curve registration methods will be introduced and discussed in section 2. Section 3 will present the actual data analysis with applications of the major curve registration methods and the different registration approaches will be compared to access the performance of different methods on wearable device data. Section 4 will make a conclusion of this study.

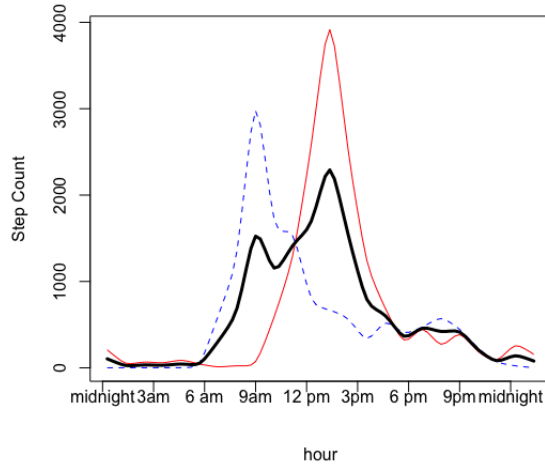


Figure 2: Actual Fitbit user step-count data of two users with similar activity pattern but misaligned step-count function. The thick black line is their cross-sectional mean.

2 Literature Review

2.1 Overview

The essence of time warping is to eliminate the so-called phase variation through aligning sample curves. To properly and efficiently align different curves, the best way to do it is through aligning curves with respect to some reference curve that usually can represent the sample curves. Such a reference curve is often referred to as the *warping template*. Intuitively speaking, curve registration is analogue to lining up a group of students where the student standing in the front of the line is often used as a target to which the rest of the students shall align. In this case, the student standing in front serves as the aligning template. One can imagine that, without a universally agreed aligning template, each student may have her own target to align to and the result would still be a misaligned procession. Based on the intuition of defining and utilizing a warping template, numerous registration methods have been proposed and the major difference between these methods lies in the type of warping template that each method uses. In general, there are three main types of template choices:

the Landmark approach uses specific point(s) on the observed curve as the warping template; the Procrustes approach uses one single curve, which is usually not an observed curve and obtained through some iterative algorithm, as the warping template; the Pairwise approach uses every single observed curve as warping templates. Here, existing warping methods will be classified based on these three template choices. Some other prominent methods that do not belong to any of these three categories will be briefly discussed at the end of this literature review. Some previous review article on functional data analysis and curve registration that provided inspiration and clarification to the composition of this literature review includes "Review of Functional Data Analysis" by Wang, et al. and "Functional Data Analysis of Amplitude and Phase Variation" by Marron, et al., both of which articles have been very helpful for this literature review [11, 12].

2.2 Landmark Approach

One of the earliest choice is the point(s) template. The method that corresponds to this warping template is the Landmark registration method proposed by Kneip and Gasser [13]. The theoretical foundation of the landmark registration method is the 1992 paper by Kneip and Gasser in which the Self-Modeling Regression Model (SEMOR) is defined to be

$$f(t) = \vartheta_1 \phi \left(\frac{t - \theta_2}{\theta_1} \right) + \vartheta_2$$

where $f(t)$ is an observed function and the $\phi(t)$ is the latent function, which is the true function to be inferred from the observed sample curves. [14]. The observed function is obtained by adding the scale parameters ϑ_1 , θ_1 and the shift parameters ϑ_2 , θ_2 . The latent function $\phi(t)$ is estimated by the information contained in the extremum of $f(t)$, which can be used to obtain information about θ_1 and θ_2 .

Based on the SEMOR model, the landmark registration method dedicates to align functional data based on their extremums (landmarks), which are then used to obtain information

about the latent function. One issue with the functional data is that the observed function are often noisy, which means that there could be a large number of false local maximum or minimums. To deal with such an issue, structural intensity, which is defined to be the number of extremums in an given interval of a pre-specified grid, is used to determine the landmarks [13]. Specifically, the pre-specified grid consists of compact subintervals of the given time domain and the number of extremums in a given interval is used to calculate the structural intensity. Formally, given an observed function r and a subinterval $I := [a_0, a_1]$ of the time domain, let $M(r') = \{\tau \in I : r'(\tau) \text{ corresponds to the maximum/minimum of } r\}$, then the structural intensity for some $t \in I$ is defined as

$$f(t) = \lim_{h \rightarrow 0} \frac{1}{2h} \mathbb{E}(|M(r')|) \cap [t - h, t + h]$$

where $|M(r')|$ represents the cardinality of the set $M(r')$. In the case of non-smooth functional data, the estimate of the structural intensity can be calculated by estimating the first and second derivatives of the data using kernel estimation. After obtaining the structural intensity, the true features (extremums) are determined and the sample curves are therefore aligned to the average location of the corresponding extremums.

Heuristically, to be able to yield a good alignment result using the Landmark registration, it is essential that the user knows exactly how many landmarks shall the functions contain. In other words, the number of extremums is determined completely based on the user's subjective judgment. While the structural intensity provides user with a reference on the credibility of the extremums, the user still needs to determine exactly how many extremums that the latent function ought to have. Hence, the alignment result is still largely depended on user judgment. Due to such nature of landmark registration, it is ideally used when people have knowledge of the common shape of the sample curves, which in turn assures the reliability of the information on the number of landmarks to be used as warping template.

2.3 Procrustes Approach

2.3.1 Dynamic Time Warping and the Procrustes Fitting Criterion

In contrast to the point-based template of landmark registration, Dynamic Time Warping (DTW) uses a complete continuous sample curve as the warping template instead of discrete extremum locations. The DTW method is first being proposed by Wang and Gasser in 1997 [15]. This method is inspired by the DTW method that engineers used to align speech signals [16]. While they are both alignment methods, the functional DTW is proposed purely from a statistical perspective and it is used exclusively to align a pair of curves. In some sense, the DTW algorithm for aligning two curves as well as the DTW algorithm for aligning multiple curves established a broader stage for developing time warping methods as it introduced the intuition of using the distance between two curves as the cost function for alignment. Differ from the landmark registration, which requires knowledge on the general shape of the sample functional data, the DTW method do not require such information and can be applied to functions where user have very few prior knowledge.

In short, the DTW algorithm works like this: The general model for DTW alignment aligns with the classical model for analyzing sample of functions, which is defined as $y_{ij} = f_i(t_{ij}) + \epsilon_{ij}$, $i = 1, \dots, n, j = 1, \dots, n_i$ where f_i is the latent functions and ϵ_{ij} are independent random variables with $\mathbb{E}[\epsilon_{ij}] = 0$ and $Var[\epsilon_{ij}] > 0$. For two smooth curves f and g , one of them is selected as the so-called warping template and the other curve is to be aligned against the template curve. Specifically, the alignment process is in fact a minimization problem that seeks to minimize the following cost function with respect to h :

$$C(f, f', g, g', h, \alpha) = \int_0^1 \left[\alpha^2 \left(\frac{f(t)}{\|f\|} - \frac{g(h(t))}{\|g\|} \right)^2 + (1 - \alpha)^2 \left(\frac{f'(t)}{\|f'\|} - \frac{g'(h(t))}{\|g'\|} \right)^2 + \phi(h'(t)) \right] dt$$

where f' and g' are the first derivatives of the functions f and g respectively and $\|f'\|$ and $\|g'\|$ are their supremum norms; h is the shift function; α is a tuning parameter that determines whether the warping shall reply more on the original functions or their first derivatives; and

$\phi(h'(t))$ is a convex penalty function. In this case, function g is to be warped against function f . It worth mentioning here that, as the formula shows, the classical DTW algorithm is only capable of aligning two functions, which is also the biggest shortcoming of this method.

Notice that the major difference between landmark registration and DTW is that landmark registration used specific points (landmarks) as the warping template whereas the DTW method considers the whole curve as a warping template. As stated above, the DTW algorithm is a dynamic programming problem that seeks to minimize the cost function $C(f, f', g, g', h, \alpha)$. The reason that the distance between the first derivatives of the two functions is included in the cost function is that the author believes that the extremums are usually the most interesting and critical feature of the functional data; Therefore, such feature shall be taken into consideration while aligning the two curves. In addition to the \mathbb{L}^2 distances, a penalty function $\phi(h'(t))$ is also added to prevent the algorithm from warping too mildly or aggressively. This process of using the \mathbb{L}^2 norm along with a penalty parameter to align two functions provided a vital intuition and a general pipeline to the later curve registration methods. In a pairwise sense, this DTW method has set the foundation for the pairwise warping method.

As we have talked before, one major weakness of the original DTW method is that it can only be used to align two curves. The author has outlined a general procedure of applying the DTW algorithm to align multiple curves in the 1997 paper and it has been explained in detail in the 1999 paper also by Wang and Gasser [17]. To deal with multiple curves, an iterative algorithm is proposed. The essence of the iterative algorithm is to estimate a best warping template instead of using an observed curve directly from the sample as the warping template. The reason that such estimation is needed here is that there is rarely a curve in the multiple curve registration case that can serve as the out-of-box warping template. Therefore, the observed curves are first aligned using the same cost function $C(f, f', g, g', h, \alpha)$ as the one used for DTW for two curves and the distance between the aligned curves and their cross-sectional mean are calculated iteratively until the improvement in the cost function

is small. It is shown that the distance between the aligned curves and their cross-sectional mean will converge eventually and thus yield the best warping result.

The spirit of this iterative algorithm of DTW for multiple curves and the Procrustes fitting criterion proposed by Ramsay and Li are similar but there are some nuances [18]. For instance, the iterative algorithm in the DTW method directly yields the warping result and the warping functions whereas the Procrustes fitting criterion used the same algorithm to find the warping template, which is the cross-sectional mean of the aligned functions in the convergent result. The initial cross-sectional mean is used as the initial "guess" of the warping template and the Procrustes fitting criterion iteratively align the observed functions and update the warping template by taking the cross-sectional mean of the aligned functions, which eventually yields the best estimate of the warping template. There are numerous warping methods that used the Procrustes fitting criterion and the registration using Fisher-Rao metric yields some of the the most prominent result among them all [10].

2.3.2 Registration Using Fisher-Rao Metric

Synchronizing curves based on a single warping template is nowadays the most popular method and it also gives some of the best warping results in terms of feature quality. As mentioned before, most single-curve warping methods are developed based on the DTW algorithm for multiple curves as well as the Procrustes fitting criterion.

The essence of the single-curve template approach is the Procrustes fitting criterion proposed by Ramsy and Li [18]. Single-Curve template approaches include methods like time warping using Fisher-Rao metric method (later referred to as the Fisher-Rao method) and curve alignment by moment [19, 10]. Among all these methods, the Fisher-Rao method proposed by Srivastava et.al. yields some of the best curve alignment result [20, 10]. The idea of defining a metric that contains information on phase and amplitude variation is first brought up by Liu and Müller in 2004. The 2011 paper by Vantini that defines a mathematical framework for the analysis of amplitude and phase variation further develops the idea

and defines a semi-metric that takes account of the phase and amplitude variation of functional data [21, 22]. Both papers set the theoretical foundation for the Fisher-Rao warping method.

In the Fisher-Rao method, instead of using the standard \mathbb{L}^2 distance and the cross-sectional mean to estimate the true latent function, Srivastava first considered two spaces \mathcal{F} , which is the space of all observed functional data, and Γ , which is the space of all possible warping functions. Srivastava believes that the true estimator of the latent signal can be obtained on the quotient space of \mathcal{F}/Γ . In this case, instead of obtaining the latent function indirectly from estimating each single warping function, which is the approach that most time warping methods take, the Fisher-Rao method searches for the true signal within a family of warping functions, which makes the method more flexible in theory and the estimator of latent function follows naturally after obtaining the warping functions.

There are, however, two major complications in the Fisher-Rao method. First, the Fisher-Rao metric is computationally complicated. For function f in the functional space \mathcal{F} and v_1, v_2 in the tangent space to \mathcal{F} at f , the Fisher-Rao metric is defined to be the inner product:

$$\langle\langle v_1, v_2 \rangle\rangle_f = \frac{1}{4} \int_0^1 v_1(t)' v_2(t)' \frac{1}{|f'(t)|} dt$$

The Fisher-Rao distance between two functions $f_1, f_2 \in \mathcal{F}$, d_{FR} , is then given as the shortest path between f_1 and f_2 under the Fisher-Rao metric. The solution to such minimization problem can only be estimated by some numerical algorithms. In other words, it is very difficult to get the precise Fisher-Rao distance between two functions. To deal with such a complication, instead of directly computing the Fisher-Rao distance between the original functions, we first map $f_i(t)$ to its Square-Root Velocity Function (SRVF) representation, which is defined to be $q_i(t) \equiv f_i'(t)/\sqrt{|f_i'(t)|}$. With such conversion, the complicated Fisher-Rao distance of the original observed functions becomes the standard \mathbb{L}^2 distance between their SRVFs (e.g. $d_{FR}(f_1, f_2) = \|q_1 - q_2\|$).

With the mapping from the observed functions to their SRVFs comes the second complication. The SRVF transformation requires information on the first derivative of the observed function, which could be troublesome for non-differentiable $f(t)$'s. There is no easy solution to such a problem and kernel estimation is employed when the first derivative of a non-differentiable functional curve is to be computed. The simplification of the calculation of the Fisher-Rao distance, on the other side, overweights the complication of estimating derivatives since the Fisher-Rao distance of the original functions becomes the standard \mathbb{L}^2 distance between their SRVFs. Therefore, converting the observed function to their SRVFs is still computationally more efficient given the SRVF representation of the sample curves can be defined.

To estimate the true warping function from a family of warping functions, the SRVF representation of a function f composite a warping function h is defined as $f \circ h = (q \circ h)(t) \cdot \sqrt{|h'(t)|}$. To make the procedure more intuitive, the concept of an orbit is introduced. The orbit of an SRVF denoted $[q]$ is the collection of SRVF q composite all possible warping functions: $[q] = \text{closure}\{(q \circ h)(t) \cdot \sqrt{|h'(t)|} | h \in \mathcal{H}\}$. In other words, the orbit $[q]$ is the set of functions that have the same amplitude variation but different phase variations. The set of all orbits is denoted \mathcal{S} and the elastic distance between different orbits of \mathcal{S} is defined to be

$$d([q_1], [q_2]) = \inf_{h \in \mathcal{H}} \|q_1 - (q_2 \circ h)(t) \cdot \sqrt{|h'(t)|}\|$$

Given a set of observed functions f_1, \dots, f_n , the Fisher-Rao method first expand the observed functions to their SRVF orbits denoted $[q_1], \dots, [q_n]$. To look for the warping template, the Fisher-Rao method first find the Karcher mean of all orbits \mathcal{S} , which can be calculated as

$$[\mu]_n = \operatorname{argmin}_{[q] \in \mathcal{S}} \sum_{i=1}^n d([q], [q_i])^2$$

Intuitively, the Karcher mean is the orbit $[\mu]_n$ that is overall the closest to all other orbits. Here, the Karcher mean $[\mu]_n$ still represents an orbit of SRVFs. Therefore, one more step

needs to be taken to find the explicit warping template function and this is the place where the concept of the center of an orbit comes into play. The center of an orbit yields the specific function that can be used as the warping template. To solve for the center of an orbit, the SRVF of observed sample curves $\{q_i\}$ are needed. For an arbitrary element μ in the orbit of Karcher mean $[\mu]_n$, we align all the observed SRVF to μ and obtain a set of warping functions by calculating

$$h_i = \operatorname{argmin}_{h \in \mathcal{H}} \|\mu - (q_i \circ h)(t) \sqrt{h'(t)}\|$$

Then the mean of all h_i is calculated to be

$$\bar{h}_n = \operatorname{argmin}_{h \in \mathcal{H}} \sum_{i=1}^n d_{FR}(h, h_i)^2$$

It follows that the warping template can be calculated as $\tilde{\mu} = (\mu \circ \bar{h}_n^{-1})(t) \cdot \sqrt{|(\bar{h}_n^{-1}(t))'|}$. After obtaining the warping template $\tilde{\mu}$, then all the observed SRVF are aligned against the template which completes the warping process.

The Fisher-Rao method produced some of the most excellent results in terms of curve alignment so the idea of curve registration using the Fisher-Rao metric has been extended to be used extensively under different scenarios. The Robust Template Estimation method used the Band-depth Median instead of the Karcher mean to calculate the warping template with the Fisher-Rao metric, which makes the warping template more robust to potential outlier functions [23]. The Fisher-Rao registration method is also applied to deterministic signal estimation and activity recognition with accelerometer data [24, 25]. The norm-preserving Fisher-Rao warping method imposed a norm-preserving constraint to the original Fisher-Rao method that guarantees the area-under-the-curve is preserved after registration [26]. Further, the Generative models for functional data proposed by Tucker, Wu, and Srivastava in 2013 also used the Fisher-Rao metric to separate the phase and amplitude variation of functional data and proposed a joint distribution to model both variations [27].

2.4 Pairwise Approach

Similar to the original DTW algorithm that aligns two curves, the pairwise curve synchronization method proposed by Tang and Müller in 2004 [28] aims to estimate the time warping function by calculating pairwise warping functions between all possible pairs of curves. There are, however, two fundamental difference between the pairwise curve synchronization method and the DTW method. First, unlike the DTW algorithm, which is dedicated to align a pair of curves, the pairwise warping method is designed to align multiple curves simultaneously and each curve will serve as the warping template to all other curves. The second difference is the perspective through which the observed functional data are studied. To be more precise, the precursor of the pairwise curve synchronization method is the 1998 paper on curve registration by Ramsay and Li and the 2004 paper by Liu and Müller, both of which brought up the idea of treating functional data as realizations of a random process which enables the use of statistical modeling in the study of functional data analysis [21, 18]. Therefore, as a result, the pairwise synchronization method assumes that the observed functional data are realizations of a bivariate stochastic process with one variate being the time warping function and another to be the amplitude function. Such a perspective is drastically different from any other previous warping methods. The idea of treating functional data as the realization of a stochastic process is a natural approach as the phase and amplitude variation are indeed considered as random quantities in time warping studies.

To be more precise, for a given time domain \mathcal{T} , the pairwise warping method assumes that the misaligned observed data (t_i, y_{ij}) can be modeled as

$$y_{ij} = Y_i(t_j) + \epsilon_{ij} = X_i(h_i^{-1}(t_j)) + \epsilon_{ij}, \quad t_j \in \mathcal{T}$$

where ϵ'_{ij} s are independent and identically distributed noisy with mean 0 and variance $\sigma^2 < \infty$ and $h_i(t)$ represents the warping function. Further, the time-synchronized random function $X_i(t)$ can also be modeled as $X_i(t) = \mu(t) + \delta Z_i(t)$ for $t \in \mathcal{T}$, where $Z_i(t)$ is

realization of random trajectory with mean 0 and finite variance. The value of δ is said to converge to 0 as the sample size increases.

Another major contribution of the pairwise curve synchronization method is its unique way of estimating the warping functions. Instead of estimating the entire smooth warping function, the pairwise warping method utilizes user-specified equal distant knots to estimate a piece-wise linear function to serve as the warping function. The advantage of this approach is that it converts a difficult problem of calculating the entire smooth curve to a simpler problem of estimating the coefficient of linear spline basis. Precisely, given p equal-distant knots (a_0, \dots, a_p) and let $\mathcal{T}_i = h(a_i)$, then the warping function can be expressed as $h(t) = \sum_{i=1}^{p+1} \mathcal{T}_i A_i(t)$ where $A_i(t)$ is the linear function in the i^{th} interval. Further, let $\Theta = \begin{bmatrix} \mathcal{T}_1 & \dots & \mathcal{T}_{p+1} \end{bmatrix}^\top$ and $A(t) = \begin{bmatrix} A_1(t) & \dots & A_{p+1}(t) \end{bmatrix}^\top$ then the warping function can be represented as

$$h(t) = \Theta^\top A(t)$$

with parameter space

$$\Omega = \left\{ \Theta \in \mathcal{T}^{p+1} \mid \Theta = (\tau_1, \dots, \tau_{p+1})^\top, 0 < \tau_1 < \dots < \tau_{p+1} \equiv T \right\}$$

To find the warping function of each curve, the pairwise warping method first obtains the warping function that aligns this particular curve to every other curve in the sample. The approach of aligning two functions is to minimize the expected value of the \mathbb{L}^2 distance between the warped function and the template function. Precisely, given two curves Y_i and Y_k the warping process is essentially a minimization problem with respect to $\Theta \in \Omega$ and cost function is defined as

$$\hat{\Theta}_{ik} = \arg \min_{\Theta \in \Omega} E \left(\int_{\mathcal{T}} \left[\{Y_i(\Theta^\top A(t)) - Y_k(t)\}^2 + \lambda \{\Theta^\top A(t) - t\}^2 \right] dt \mid Y_i, Y_k \right)$$

which yields the preliminary estimate of the coefficients of the linear spline bases that aligns

function Y_i against Y_k . In this case, the same minimization problem is then solved to align the specific function Y_i against all other functions in the sample. Further, to avoid aggressive warping, a penalty parameter is also included in the minimization problem that avoids the warping function from severely distorting the shape of the original curve. It follows that the estimator of the inversed true warping function is then the arithmetic mean of the pairwise warping functions

$$\hat{h}_k^{-1}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\Theta}_{ik} \cdot A(t)$$

and the inverse of the estimated inversed warping function is guaranteed to exist since $h_k(t)$ is defined to be a diffeomorphism. Thus, the synchronized function can be calculated as $Y_{\text{aligned}}^k = Y_k \left(\hat{h}_k \right)$.

This is notably different from the methods that do not treat functional data as realizations of a random process. Take the previous landmark registration and DTW as an example, the warping function is directly estimated by aligning the curves with a previously defined template. In the pairwise case, the warping function is not directly obtained in the aligning process. Instead, numerous estimates of warping function of each curve is calculated and their cross-sectional mean is then used as the final estimate of the warping function. Such a way of estimating warping function is theoretically sound because of the fact that the sample curves are treated as realization of a random process. Otherwise, taking the cross-sectional mean of all the pairwise warping function would not make sense as the curve shall only be aligned to the potential latent function.

2.5 Other Time Warping Methods

The stochastic perspective in functional data analysis brought up by Ramsay and Li allowed for more probabilistic approaches such as the Adaptive Variational Bayes approach proposed by Earls and Hooker that modeled functional data as a stochastic process and used the novel Adaptive Variational Bayes approach to estimate the true latent function [29]. Similarly, the

Registration for Exponential Family Functional data method proposed by Wrobel et. al. also assumes that the latent true signal, which is regarded as an "unobserved smooth probability curve", follows a Gaussian process [30]. In this case, the parameter of the latent function is then estimated using the Expectation-Maximization algorithm and a similar variational Bayes approach.

The same idea of using piecewise linear warping functions to simplify the complexity of estimating the complete smooth warping functions is also proposed by Gervini and Gasser in the same year as the pairwise warping method [31]. Such an approach is then further studied by Chakraborty and Panaretos and a tuning-free method is proposed using the fact that the piecewise linear warping function can be obtained by solely estimating the coefficients of the linear spline basis [32]. Originally, a penalty parameter is needed in the pairwise warping method to penalize shape distortion. In the tuning-free method, local variation, which is the amplitude variation on each partition of the time domain, is used to obtain an estimation of the warping function. Time warping is a process that aim to eliminate the phase variation and leave the amplitude variation unchanged. Therefore, the amplitude variation of each partition will only be redistributed after warping. Therefore, the relationship between warping and local amplitude variation is predictable, which is then used to estimate the warping function while guarantees that the shape of the original function is not distorted by controlling its local variations.

Another interesting curve registration method is the Registration and Fitting method proposed by Kneip and Ramsay [33]. In the previous articles, the motivation of curve registration is to eliminate as much phase variation as possible so that statistical tools that could be interfered with phase variation such as Functional Principal Component Analysis can be employed. In other words, curve registration has historically been regarded as a pre-processing step for later data analysis. This method proposed a brand new perspective on approach curve registration. The question that this method raised is that how well shall the warping result be and how shall analysts determine the statistical interpretability of the

warping result and prevent over-registering and under-registering. In other words, the fact that all later analysis results are dependent on the time warping result urges researchers to think more carefully with how to preserve the statistical information contained in the original data while registering them. The method proposed here is to iteratively conduct functional principal component analysis and the register to the Karhunen-Loève Decomposition of the observed data $X_i(t)$ (refer to (1)). This expansion sets the foundation for Functional Principal Component Analysis and also the Registration and Fitting method, which is based on the method of functional principal component analysis. Unlike the Procrustes fitting criterion that most single-curve template warping methods used, the Registration and Fitting method iteratively calculate the estimation of the latent function by register the function from the previous iteration to the function given by the Karhunen-Loève expansion of the current iteration. The increment of the warping function is the criterion used to determine the convergence of the iterative algorithm. With such a method, it is guaranteed that the phase variation between functions is reduced and the warping result is still statistically interpretable.

3 Data Analysis

3.1 Introduction

The data that are used for this data analysis are obtained from a survey distributed through the Amazon Mechanical Turk that consists of recording of 34 users' health data including step count, calorie burned, sleep quality, etc. from March 12th 2016 to May 12th 2016. The specific feature of this data that is being studied here is the step count data recorded every hour through the two months period. Explicitly, the hourly step count data is used to examine the result of three major time warping methods: Landmark Registration, which is one of the earliest time warping method, Pairwise Curve Synchronization, which used a unique pairwise alignment approach, and Curve Registration with Fisher-Rao Metric, which

yields the most prominent warping result in all template warping methods.

Since the dataset consists of repeated measurements of step-count data on each of the weekdays, obtaining a representative curve of users' step-count data for each weekday is the crucial step toward a meaningful data analysis result. In this data analysis, two approaches of implementing the warping methods are attempted in this data analysis. The first approach is to naively take the mean of each user's step count on each day of the week so that an approximation of how the user's average step count data would look like on different week days is obtained. This approach is later referred to as the "1-step warping" approach. One flaw that the 1-step warping approach has is that the repeated measurement of user step-count data could also be misaligned resulting in the cross-sectional mean being unrepresentative to the actual activity pattern that the user may have. This concern leads to the second approach. In the second approach, instead of naively taking the mean of the repeated measurements of user step-count on each day of the week, the weekday step-count data of each user is first registered to ensure that the cross-sectional mean we obtained is representative. After the first registration step, a second registration is then employed on the mean of the registered user step count data to obtain a representative curve that conveys information on the overall user activity pattern on each weekday.

To access the warping result, FPCA is used as a measurement since the main purpose of time warping is to eliminate phase variation so that methods that only analyze amplitude variation like FPCA can be more accurately interpreted. Specifically, the warping results are measured with the cumulative fraction of variance explained (FVE) by the leading eigenfunctions by conducting FPCA after the curve registration algorithm. Generally, better warping result will be reflected by the leading eigenfunctions explaining more variation, which indicates that amplitude variation is more dominant over phase variation. It is also important to clarify that FPCA is not the only measurement that is considered in this data analysis, other factors like computing efficiency of warping algorithms and feature quality after warping are also considered to ensure that the warping results are not overtly artificial and that they

have some statistical interpretability.

To better compare the three major warping methods, two plots consist of the warping result and the warping functions are provided in Figure 3 through 6:

3.2 Landmark Registration

3.2.1 1-Step Landmark Registration

First of all, the number of landmarks is set to be 2 based on visual inspection of the original data that most individual curve has two peaks. From the warping functions of 1-step curve registration shown in Figure 3, we can see that some warping functions have very steep or flat slope, which indicates that the corresponding curves are severely distorted on the time scale. Such aggressive warping occurred due to the two "peaks" of the curve being too close to each other. With that being said, such issue is also easily resolvable by manual adjustment of the landmarks. Therefore, the landmark registration method is very flexible and it is capable of conducting adequate registration when the landmarks are properly set.

There are, however, two major issues with the landmark registration. The first issue is the low efficiency of the algorithm limits landmark registration from being used on large datasets. In this study, landmark registration is used 7 times to align the step-count functions of each day of the week. To conduct landmark registration, 7 sets of landmarks is manually extracted from the data. It takes around 34 seconds for the landmark registration algorithm to finish aligning 34 curves for each day. Thus, to handle 7 datasets each with only 34 curves, the landmark registration algorithm will take around 4 minutes. Notice that the dataset of this study is quite small and it is implausible to manually adjust the landmarks for large dataset. Therefore, although the aggressive warping can be tuned down by manually identifying the landmarks, it will become a severe issue with large datasets when landmarks can only be identified by certain algorithms.

To access the alignment result, FPCA is conducted and here are the average of cumulative fraction of variance explained by the first 5 eigenfunctions for data from Monday through

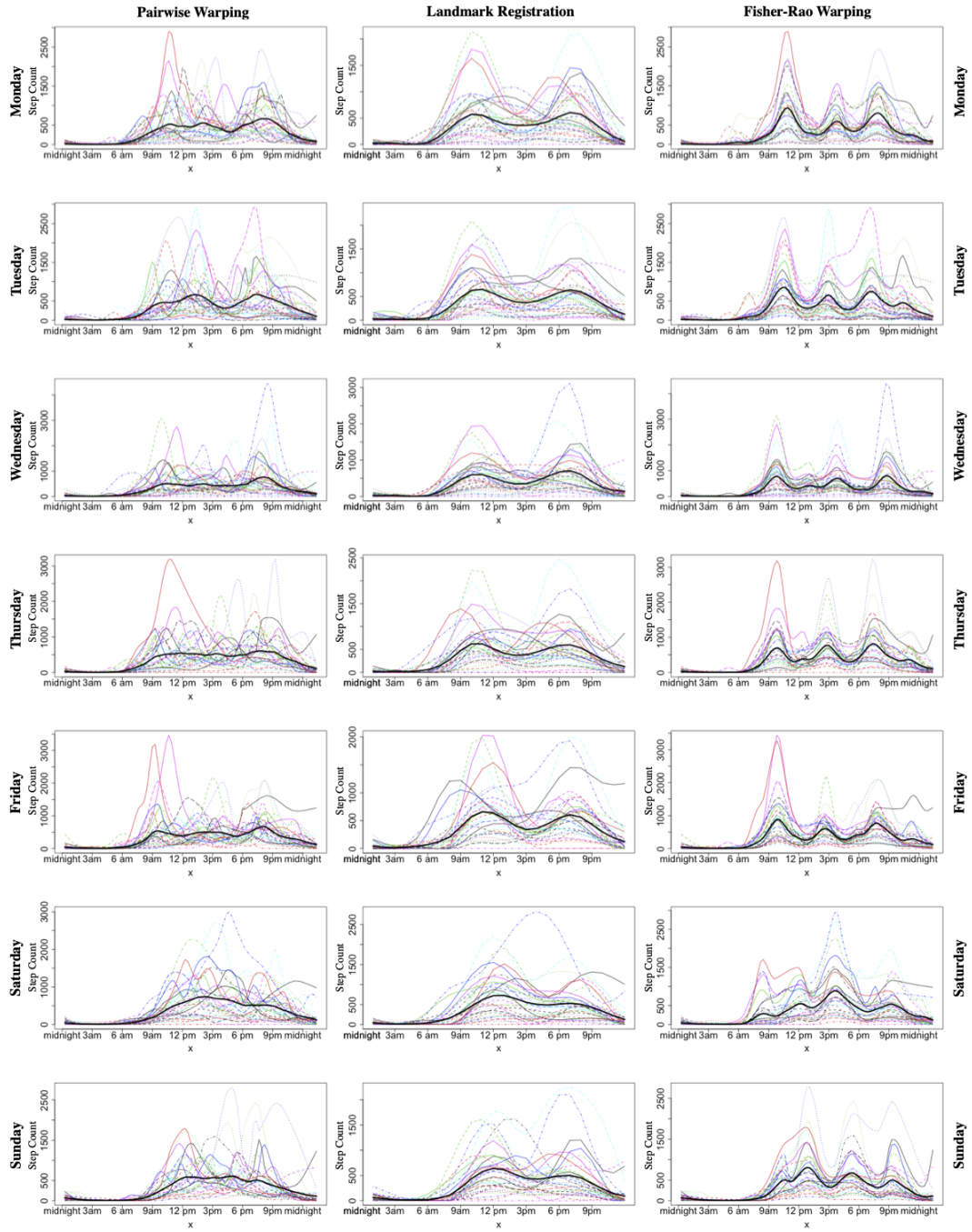


Figure 3: Warping result of 1-Step warping approach. Left column: Pairwise Warping result, Middle column: Landmark Registration result, Right column: Fisher-Rao Warping result.

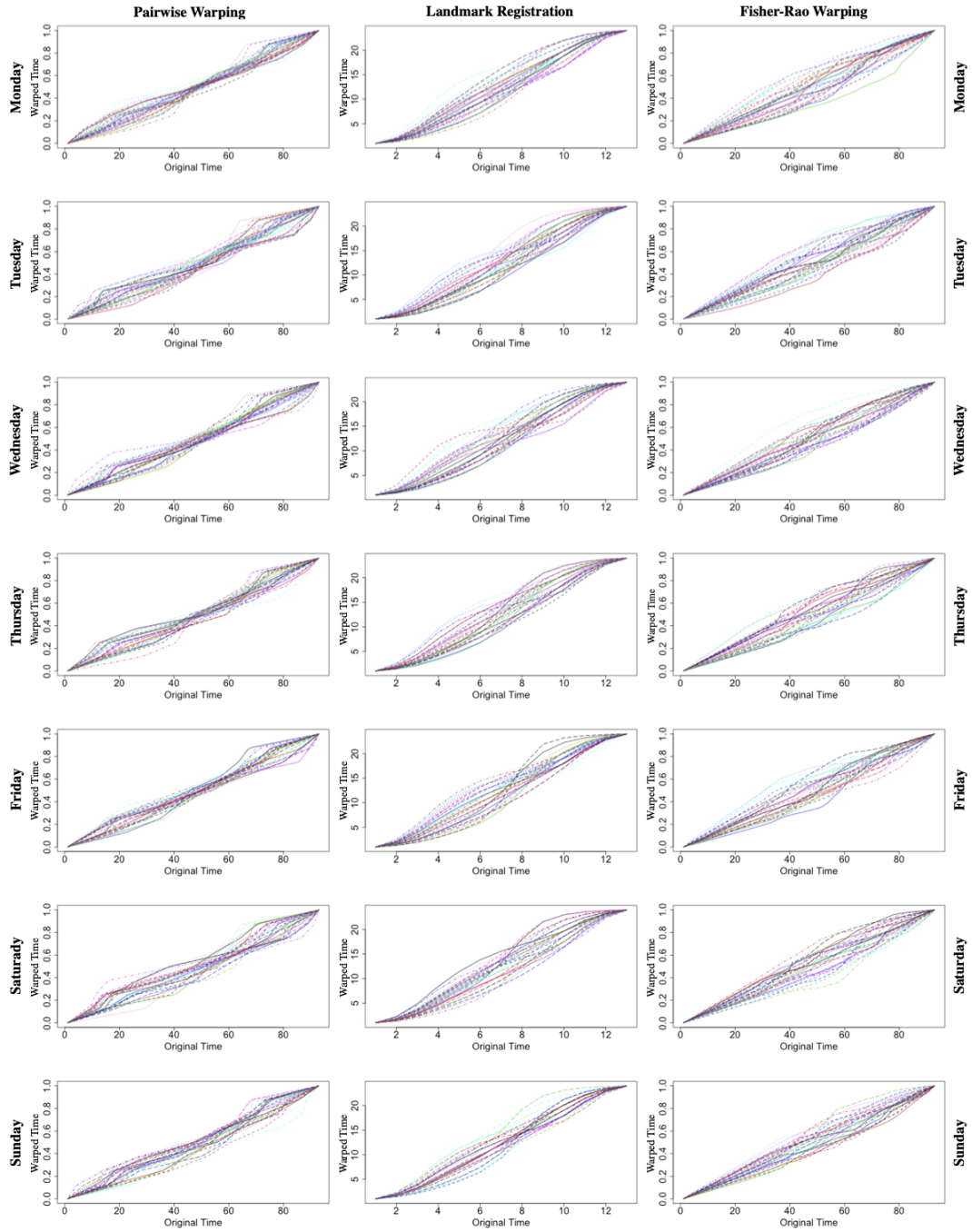


Figure 4: Warping functions of 1-Step warping approach. Left column: Pairwise Warping function, Middle column: Landmark Registration warping function, Right column: Fisher-Rao Warping function.

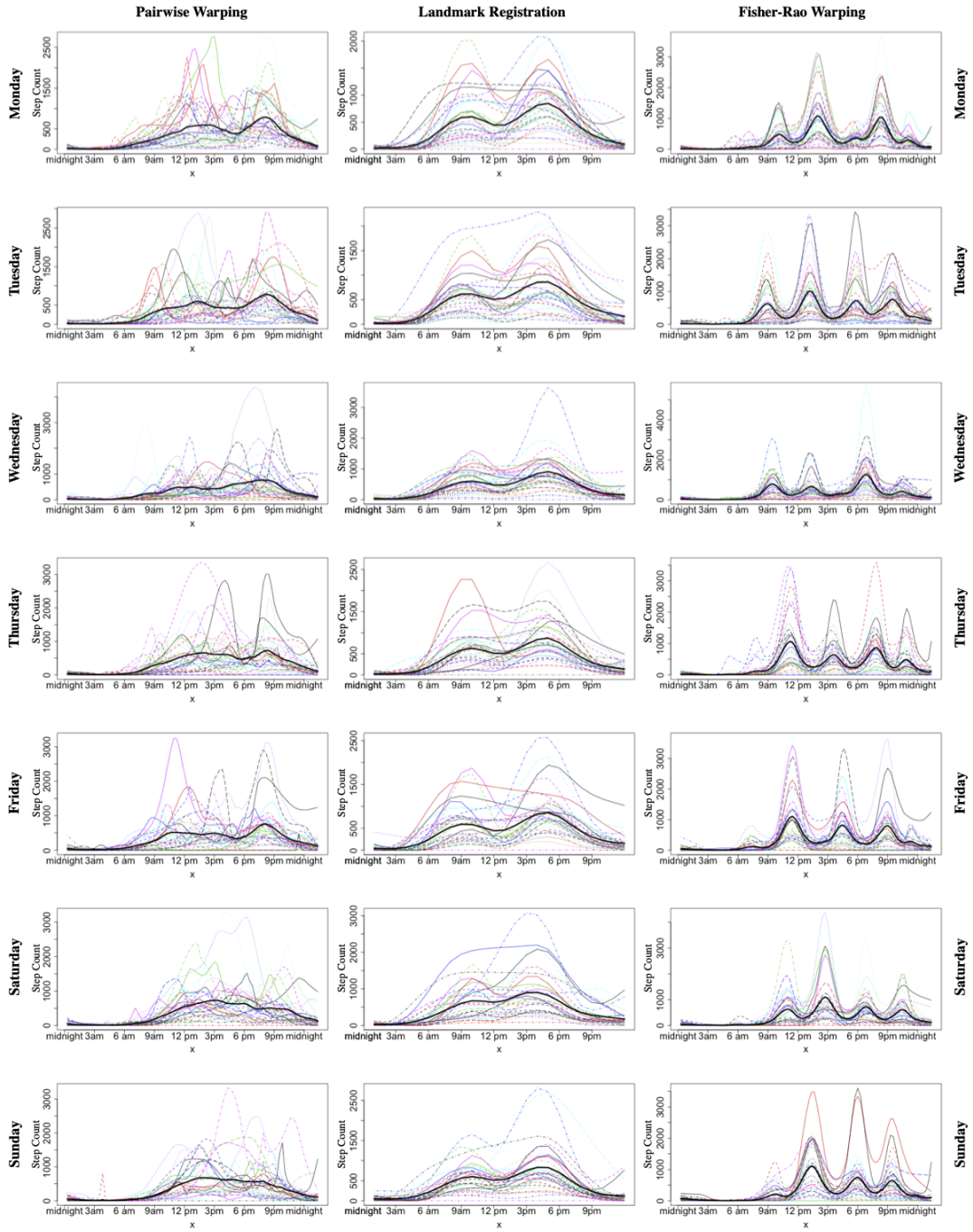


Figure 5: Warping result of 2-Step warping approach. Left column: Pairwise Warping result, Middle column: Landmark Registration result, Right column: Fisher-Rao Warping result.

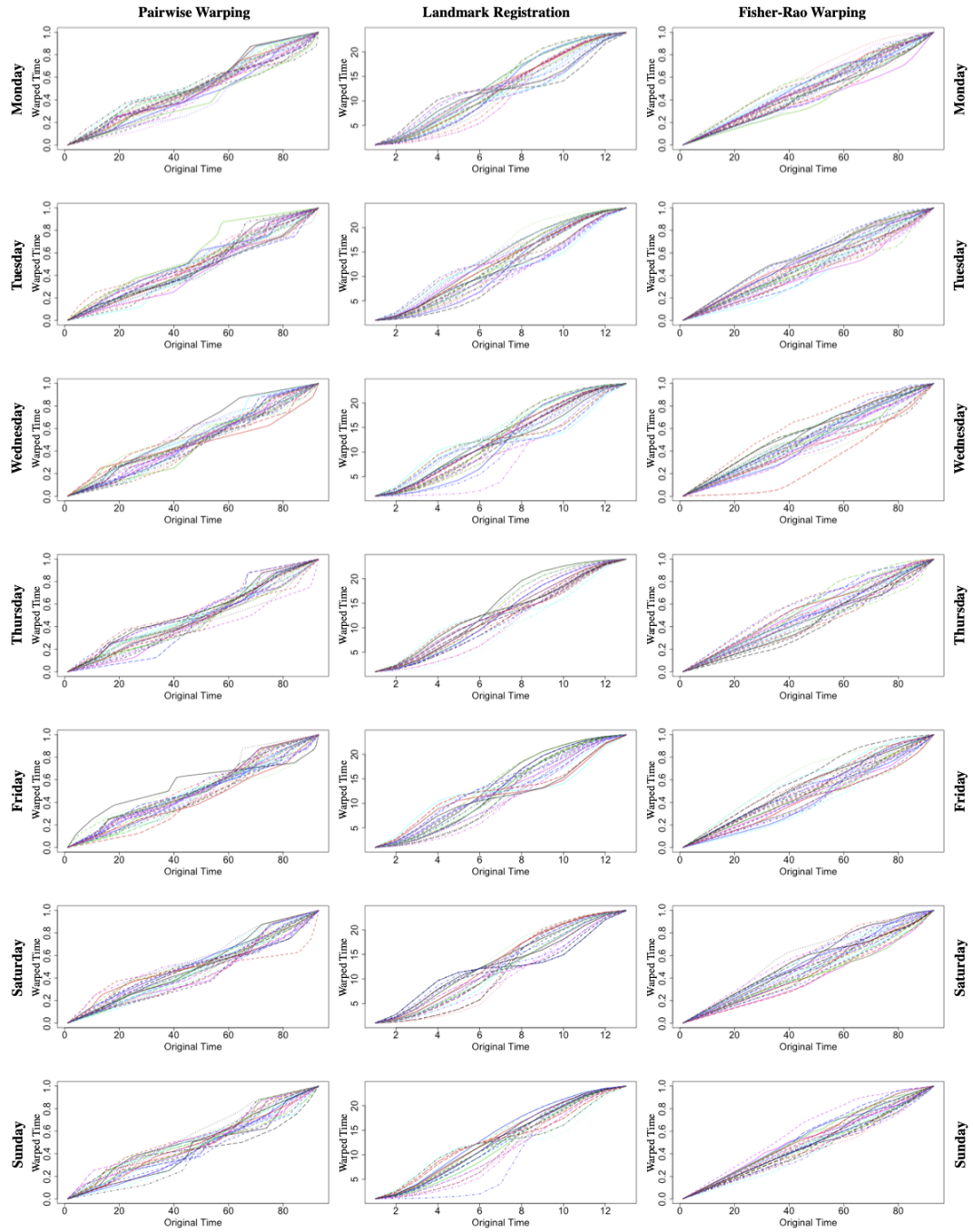


Figure 6: Warping functions of 2-Step warping approach. Left column: Pairwise Warping function, Middle column: Landmark Registration warping function, Right column: Fisher-Rao Warping function.

Sunday shown in Table 2.

Number of Eigenfunctions:	1	2	3	4	5
Monday	36.77	62.91	72.55	80.41	86.79
Tuesday	34.99	59.47	70.99	80.07	85.94614
Wednesday	45.25	61.76	74.21	83.72	88.57
Thursday	34.52	62.01	72.82	80.59	85.95
Friday	33.15	59.20	71.22	79.46	85.06
Saturday	55.32	75.74	86.51	90.90	93.84
Sunday	46.98	68.39	78.07	86.46	92.01

Table 1: FVE of the leading FPCs of Unregistered Data

Number of Eigenfunctions:	1	2	3	4	5
Monday	59.77	84.55	94.29	97.12	98.60
Tuesday	61.81	89.01	92.96	96.14	98.00
Wednesday	61.99	83.92	90.29	94.50	97.97
Thursday	64.27	85.85	91.53	95.95	98.37
Friday	60.04	81.86	89.14	95.65	98.12
Saturday	76.51	87.99	95.27	97.48	98.85
Sunday	73.85	88.57	93.69	96.24	98.20

Table 2: FVE of first 5 eigenfunctions of data from Monday to Sunday

Overall, the cumulative FVE of the first 3 eigenfunctions is around 90% and the cumulative FVE of the first 5 eigenfunctions is around 98%. To better study the effect of curve registration, we compare the FVE of the leading FPCs of the registered data with the FVE of the leading FPCs of the unregistered data, which is listed in Table 1. The improvement in the proportion of variance explained by the leading eigenfunctions reveals that, after employing landmark registration, the data are aligned and most phase variations are eliminated. Plus, the landmark registration yields the best alignment for data on the weekends, which may suggest that the latent step count function of users are more similar on the weekends compared to the weekdays. In addition to good cumulative FVE results, the mean function of step count data after landmark registration also shows very obvious features. Further, compare the warped function to the original data, we see that the warping is not very artificial. Therefore, the 1-step landmark registration would be a good candidate to be applied

to this data.

3.2.2 2-Step Landmark Registration

With 2-step landmark registration approach, it is implausible to find the landmarks manually as a total of $34 \times 7 = 238$ set of landmarks need to be performed. Therefore, the landmarks here are found by detecting the maximum step count (which is regarded as the peak of the step count function) before and after 12pm. Specifically, the boundary points (0:00 and 23:00) are first removed so that the domain of the curve is not changed after registration. The time is then separated into two intervals— one for the morning (before 12pm) where the first peak usually occur and the second for the afternoon (after 12pm) where the other peak often occur. The time corresponding to the maximum value of each interval is then used as the landmarks. This approach is simple and efficient but it has one major problem, which demonstrates the potential issue of landmark registration— Since the landmarks are found completely by the algorithm, some landmarks are very close to each other (e.g. 11:00 and 13:00) which makes the warping very aggressive and the aggressive warping is displayed in the warping functions (steep and flatten slopes). To access the warping result of the 2-step landmark registration method, we first look at the cumulative FVE from FPCA.

Number of Eigenfunctions:	1	2	3	4	5
Monday	68.31	91.18	95.12	97.67	99.24
Tuesday	71.39	94.64	97.24	98.69	99.40
Wednesday	71.96	90.76	94.69	97.70	99.21
Thursday	66.94	93.27	97.41	98.39	99.15
Friday	65.76	89.77	94.38	97.30	99.17
Saturday	79.76	92.56	96.33	98.74	99.30
Sunday	78.98	94.25	97.04	98.40	99.28

Table 3: FVE of first 5 eigenfunctions of data from Monday to Sunday

From the cumulative FVE of the leading eigenfunctions, we can see that the 2-Step landmark registration approach indeed give a better result compare to the 1-Step warping. The FVE of the first three eigenfunctions increased from around 90% to around 95%. Plus,

to capture 90% of the variance, the 2-step landmark registration method only require two eigenfunctions, which further reduced the dimension of the data from 3, which is the number of eigenfunction needed to explain 90% of variance in the 1-step landmark registration approach, to 2, which is the number of eigenfunction needed using the 2-step landmark registration method. The debate, however, now becomes whether it is worth the time (low efficiency) and the distortion (bad landmarks) to reduce the data by only one dimension. To be more precise, the run time for 2-step landmark registration is 37 minutes, which has increased 33 minutes compare to the 1-step landmark registration method (4 minutes).

To conclude, the 1-step landmark registration, although needs a little bit more eigenfunctions to catch an adequate amount of variance, is a better approach comparing to the 2-step algorithm as the 2-step approach not only takes a lot longer time to run but it also distorts the original shape of the data more severely. With such heavy distortion, the increase in the cumulative fraction of variance explained is negligible. Further, in large dataset where the landmark must be detected using an algorithm, it is vital to determine whether the shape distortion would be severe. The warping in the 1-step landmark registration approach was aggressive and it was tuned down manually. Manually modification, however, is not applicable in large dataset and the algorithm used to detect landmarks will need to be designed more carefully to take care of the potential shape distortion issue.

3.3 Pairwise Curve Synchronization

3.3.1 1-Step Pairwise Curve Synchronization

First, from the aligned functions shown in Figure 3, we can already see that the latent feature is not displayed as clearly as that of the landmark registration. Part of the reason that the latent function yielded by the pairwise warping is mushier is that the pairwise warping algorithm warps less aggressively comparing to the landmark registration method. The warping functions in Figure 4 show that there is no severe shape distortion in any curve whereas landmark registration tends to warp some curves more aggressively. In addition, the pairwise

warping algorithm is the most efficient method comparing to the other two warping methods used (landmark registration, warping with Fisher-Rao metric). The pairwise warping algorithm only take 8 seconds to finish warping 34 curves, which is significantly faster comparing to the landmark registration (34 seconds) and the Fisher-Rao method (30 seconds). To be more precise, pairwise warping method takes around 1 minute to warp the 34 curves for all 7 days of the week and the run time for Landmark registration and Fisher-Rao method on the same 7 datasets are 4 minutes and 3.5 minutes respectively.

On the other hand, since pairwise warping method deals with a combinatorial number of warpings to be done during the algorithm, the run time of this method might vary significantly on larger dataset. Further, from the plot we can see that the 1-step pairwise warping result is not very meaningful. The mean function of the after-warping curves did not show obvious features as the landmark registration and the improvement between the original arithmetic mean and the warped mean is not significant. Therefore, we quantify the result using FPCA results:

Number of Eigenfunctions	1	2	3	4	5
Monday	41.94	68.83	78.66	85.24	89.03
Tuesday	49.27	70.01	83.01	87.44	91.38
Wednesday	48.91	68.35	79.78	86.90	90.07
Thursday	42.37	69.96	79.19	86.54	90.25
Friday	40.92	65.80	75.77	82.93	89.54
Saturday	74.03	84.51	89.49	93.29	96.19
Sunday	54.24	75.98	85.35	90.10	93.32

Table 4: FVE of first 5 eigenfunctions of data from Monday to Sunday

From the table, we see that an interesting phenomenon of the pairwise warping is that it performs significantly better on Saturday data. From prior section, we see that the landmark registration method performs well on both Saturday and Sunday. Hence, combine the result from landmark registration and pairwise warping, we see that the latent feature of Saturday data is more present than other days and we expected the later methods to perform the best on Saturday data.

Except for Saturday data, the 1-step warping method did not work well in general. From the FPCA result on the pairwise warped data, we can see that the first three eigenfunctions only capture around 80% of the variation, which is less than that of the landmark registration whose first three eigenfunctions captured 95% of the variance. In addition, if we compare the FVE of the 5 leading FPCs with that of Table 1, we can see that the 1-step pairwise warping did not actually help the FPCA to perform better in terms of dimension reduction. Therefore, we proceed to the 2-step pairwise warping method to see if there is any improvement in the warping result.

3.3.2 2-Step Pairwise Curve Synchronization

By looking at the path plot for 2-Step Pairwise Curve Synchronization in Figure 5, we can already see that the feature quality has improved a lot. Further, the whole 2-step warping process for pairwise warping will only take around 23 seconds, which is still faster than the 1-step warping approach for both landmark registration and warping with Fisher-Rao metric. There are, however, a few curves that are more severely distorted but it is not as severe as the distortion caused by landmark registration. The cumulative FVE of the leading eigenfunctions from FPCA is

Number of Eigenfunctions	1	2	3	4	5
Monday	50.83	74.94	83.86	88.52	91.78
Tuesday	42.35	69.81	78.30	84.55	89.87
Wednesday	57.09	68.43	77.52	85.42	92.40
Thursday	53.68	74.77	86.97	91.45	93.81
Friday	47.44	80.01	88.13	91.68	94.21
Saturday	66.80	78.94	88.52	92.64	95.65
Sunday	58.64	77.68	88.13	93.07	95.10

Table 5: FVE of first 5 eigenfunctions of data from Monday to Sunday

Here, we see a large improvement both visually and quantitatively. The after-warping mean function now shows more obvious features that the curves from Monday to Friday generally have two peak and the first three eigenfunctions can explain around 88% of the

total variance in general, which is a large improvement compare to the FPCA result on the unregistered data. The result that the latent function has two peaks matches the warping result from the landmark registration method, which indicates that the choice of setting two landmarks makes sense.

One interesting result from the 2-step pairwise warping is that, unlike the previous methods, the 2-step pairwise warping algorithm did not have any significantly better performance on Saturday data. On the other hand, the algorithm did not perform well on Tuesday and Wednesday data as the cumulative FVE of the first three eigenfunctions are significantly lower than the other data. Yet, considering the less aggressive warping approach of the pairwise warping method and a good result of most data using the 2-step approach, the pairwise warping method would be a good warping option.

3.4 Warping with Fisher-Rao Metric

3.4.1 1-Step Fisher-Rao Warping

The SRVF warping with Fisher-Rao metric by far has the best feature quality after warping. In addition, the warping function shows that there is no severe distortion done to the original curve. Unlike the previous landmark registration method and pairwise warping method, which shows that the latent function has two peaks, the warping result of the Fisher-Rao method shows that the latent function generally has three peaks. By visual inspection, there are indeed some curves that have three peaks. Therefore, it needs to be further studied to examine the shape of the actual latent function. For a rudimentary examination, then landmark registration method with three landmarks is conducted and the result in Figure 7 still shows that the true latent function shall have two peaks instead of three.

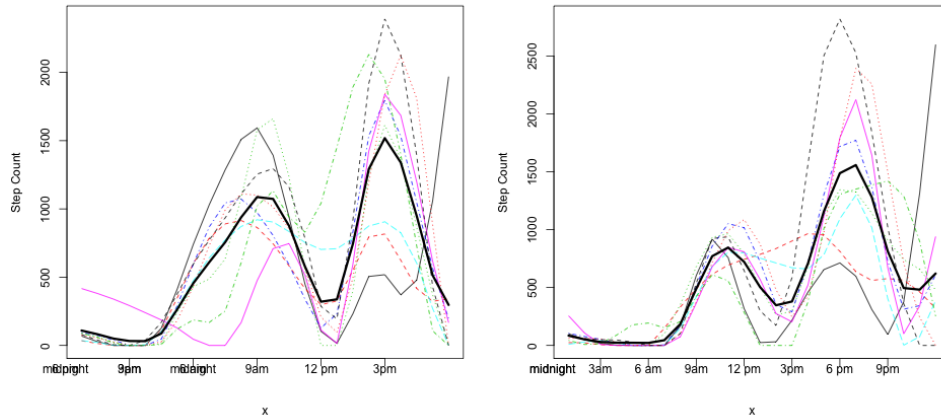


Figure 7: 1-Step Landmark Registration Result for Step-Count Data on Monday with 3 landmarks. The thick black line represents the arithmetic mean. From right to left are LM registration with 3 landmarks and LM registration with 2 landmarks.

Therefore, different algorithm yields different latent function in this study. The significance of this issue is that the corresponding FPCA will also yield drastically different result and we will have different interpretation on the same data. Therefore, the reason that caused such difference will need to be studied more closely and carefully. Due to the time limit of this thesis, such difference is not studied but it will be further studied in future projects.

The efficiency of the Fisher-Rao method is better than the landmark registration algorithm but worse than the pairwise warping algorithm. With parallel computing, the efficiency of the Fisher-Rao method is close to that of the pairwise warping method. Although the amount of iterations of computing the Karcher mean can be adjusted, to get a better approximation, this algorithm is still computationally demanding. It takes around 30 seconds to warp 34 curves with parallel computing which implies that the algorithm will need around 3.5 minutes to warp all datasets from 7 weekdays. After conducting FPCA on the warped dataset, the cumulative FVE of the leading eigenfunctions are

Number of Eigenfunctions	1	2	3	4	5
Monday	59.59	82.05	89.34	93.98	95.97
Tuesday	55.61	78.55	88.06	93.62	96.20
Wednesday	53.96	78.35	89.83	94.23	97.22
Thursday	51.77	79.97	92.38	95.51	97.05
Friday	55.72	80.72	88.80	94.04	96.26
Saturday	69.24	85.25	92.31	95.94	97.89
Sunday	69.14	83.71	92.65	95.92	97.18

Table 6: FVE of first 5 eigenfunctions of data from Monday to Sunday

This result is slightly worse than that of the landmark registration. Yet, considering the more efficient algorithm, the Fisher-Rao method is the most balanced method in terms of efficiency and warping result. Similar to the landmark registration result, the warping result of the Fisher-Rao warping method is more superior for weekends, which suggests that people tend to have similar activity pattern on weekends comparing to weekdays.

3.4.2 2-Step Fisher-Rao Warping

Analogue to the 1-Step Fisher-Rao warping result. The features are very well presented but it is different from that of the other two warping methods, which will need further study. The Fisher-Rao warping method is better to be used with a 1-Step warping approach since the result improvement is minor comparing the 1-step and 2-step Fisher-Rao warping results and that the 2-step warping approach with Fisher-Rao warping takes about 200 seconds (3 minutes 20 seconds) to warp data from a single day which increased its run time of warping all datasets increased from 3.5 minutes to 23.3 minutes. Further, the FPCA conducted on the warped data gives the following cumulative FVE of the leading eigenfunctions is

Number of Eigenfunctions	1	2	3	4	5
Monday	64.27	82.92	92.13	94.63	96.68
Tuesday	49.23	68.28	86.34	94.89	97.09
Wednesday	57.48	81.76	90.69	95.21	96.63
Thursday	57.95	77.55	87.56	93.54	95.69
Friday	54.89	77.59	90.51	94.70	97.26
Saturday	60.57	80.35	90.03	95.94	97.50
Sunday	68.66	81.76	89.89	95.56	97.03

Table 7: FVE of first 5 eigenfunctions of data from Monday to Sunday

As expected, since the result of the 1-step Fisher-Rao warping is already prominent, the 2-step warping produced almost identical FPCA result. Therefore, giving the the 2-step Fisher-Rao warping is computationally expensive. It is better to use the 1-step Fisher-Rao warping approach.

3.5 Discussion of Data Analysis Result

Besides discussing the Fitbit data from the perspective of curve registration, we also studied the real-world implication that we main obtain from the step-count data. The 2-step pairwise warping method is thought to be the most balanced registration method– the warping result is good and the warping function shows no sign of severe distortion of the original curves. Hence, the FPCA result of the 2-step pairwise warping method is further studied.

We first look at the cross-sectional mean of the step-count data, which shall be a good representation of the latent function of each weekday. By looking at the mean function, it is obvious that the users' activity pattern on the weekdays is different from that of the weekends.

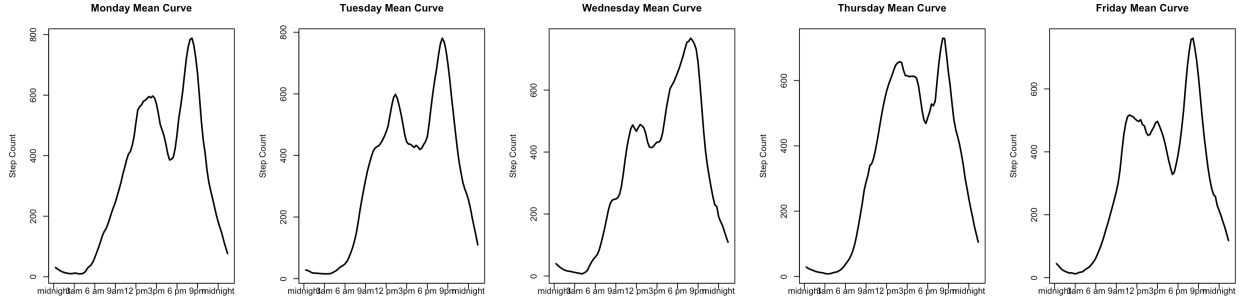


Figure 8: Cross-sectional mean function of user step-count from Monday to Friday.

The cross-sectional mean shows that the average user generally have two activity peaks in the weekdays, which could correspond to the step people take to go to work as well as the step take to go back home. Further, we see that the second peak is generally larger than the first one indicating the most user would be more active or going to more places after work.

The leading eigenfunctions, which correspond to the main modes of variation is also of interest in FDA. The 3 leading eigenfunctions of user step-count data on weekdays are displayed in Figure 9. The first eigenfunction of Monday, Thursday, and Friday data shows that many people's activity intensity differs the most at the two peaks. The first eigenfunction of Tuesday and Wednesday data show that some people's activity intensity differ only at the second peak. The first eigenfunction of Thursday data show that some users' activity intensity differ the most at the first peak. The second eigenfunction of Monday and Thursday data show that some user are only active at the two peaks and extremely inactive in between the two peaks, which may reflect the work of these users is stationary. The same eigenfunction could also indicate that some people have a one-peak activity pattern during these days. The second eigenfunction of Tuesday and Wednesday data shows that some user are more active in the first peak and less active in the second peak suggesting a different life style than what the main stream activity pattern that the mean function has suggested. The second eigenfunction of Friday data reveals that some user are significantly more active in the middle of the day, suggesting that some people may get off work earlier on Friday and tend to move around more during the day of Friday. The third eigenfunction also contain

valuable information. The third eigenfunction of Monday, Tuesday, Wednesday, and Friday data show that a small group of people tend to be more active at one of the two peaks and less so in the other peak, which also suggest a different living habit. The third eigenfunction of Wednesday data suggest that some user are less active in the two peaks and more active in the middle of the day. Hence, from the eigenfunctions, we have seen several variations of life style from the main stream pattern suggested by the mean function:

- Some people are more active in the first peaks instead of in the second peak.
- Some people are more active in the second peaks instead of in the first peak.
- Some people are only active in the two peaks and very stationary during the day.
- Some people are less active in both of the two peaks and very active in the middle of the day.
- Some people may have a one-peak activity pattern on Monday and Thursday.

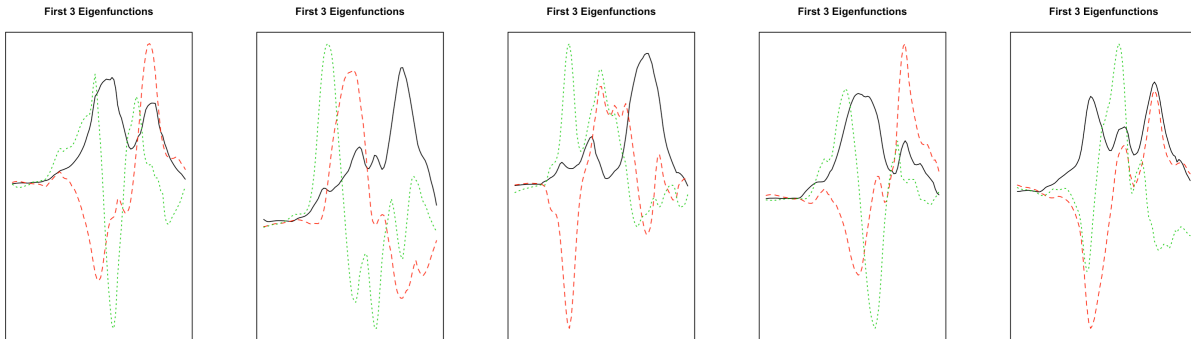


Figure 9: 3 Leading eigenfunctions of weekday user step-count data. From left to right: Monday-Friday. The black, green, and red line correspond to the first, second, and third eigenfunction.

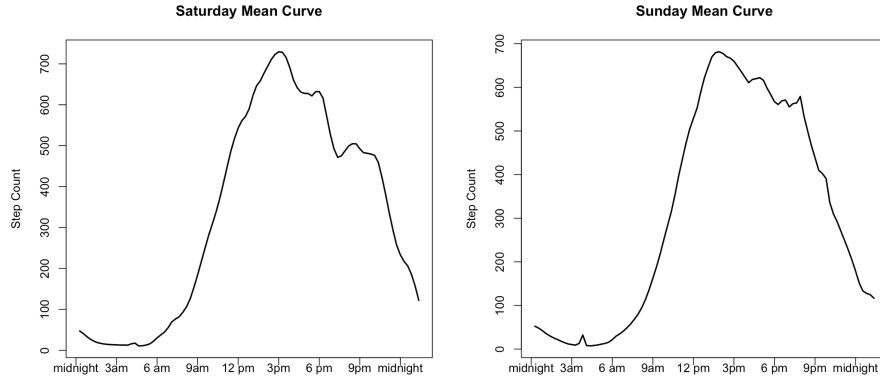


Figure 10: Cross-sectional mean function of user step-count on Saturday and Sunday

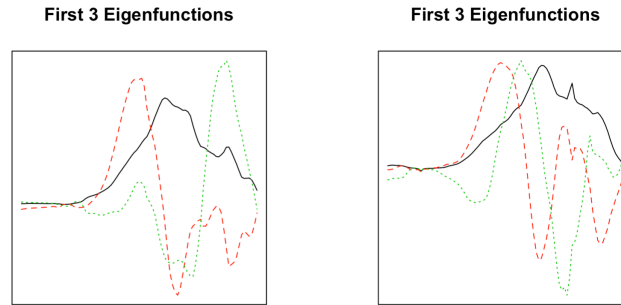


Figure 11: 3 Leading eigenfunctions of weekday user step-count data. From left to right: Saturday and Sunday. The black, green, and red line correspond to the first, second, and third eigenfunction.

Now we diversify our attention to the cross-sectional mean of the weekends data, which is displayed in Figure 10. Differs from the cross-sectional mean function of the weekdays, the cross-sectional mean function of user step-count on the weekends suggest that most user only have one peak, which occurs roughly in the middle of the day. In addition, the step-count peak sustains longer period of time on Sunday comparing to that of Saturday, which indicates that most user tend to be more active on Sunday.

Besides the cross-sectional mean function, the 3 leading eigenfunctions of the Saturday and Sunday step count data is displayed in Figure 11. As suggested by the first eigenfunction, many user are more or less active during the peak activity hours on Saturday and Sunday. The second eigenfunction of Saturday data indicates that some user are less active during the day and more active in the night or more active in the day and less active at night.

The second eigenfunction of Sunday data reveals that some people have a two-peak activity pattern as what most people do during the weekdays. Besides, the second eigenfunction of Sunday data could also indicate that some people have a narrower peak activity period on Sunday compare to most other users. Last but not the least, the third eigenfunction of both Saturday and Sunday data reveal that some user still have a two-peak activity pattern during weekends or the eigenfunctions could also indicate that some user are very inactive during the weekends and their peak activity hour is significantly shorter than that of most other users. Hence, from the eigenfunctions of the weekend step-count data, we may draw the following conclusions:

- Some people are significantly more active during the peak hours.
- Some people are significantly less active during the peak hours.
- Some people's peak activity hour is much shorter than that of most other users.
- Some people's peak activity hour is longer than that of most other users.
- Some people are more active in Saturday night where as some people are less so.
- Some people still have a two-peak activity pattern even during the weekends.

3.6 Conclusion of Data Analysis

Based on the result obtained using the three major warping methods, we see that there are two interpretations to the warped data. Therefore, the conclusion will be divided into two sections. The first section will include interpretation of the data warped by the landmark registration and the pairwise curve synchronization as they display similar latent functions. The second section will discuss the data warped by the Fisher-Rao warping method as it shows different latent function.

3.6.1 Landmark Registration and Pairwise Curve Synchronization

According to Figure 3 and Figure 5, the warped functional data using landmark registration and pairwise warping method shows that the latent function has one peak on Saturdays and two peaks on the rest of the week days. This indicates that there is generally one specific time of the day on Saturday where people walk the most and there are two times of the day on the rest of the weekdays where people are the most active. Furthermore, to examine the major modes of variations, FPCA and the mean curves along with the three leading eigenfunctions of each data is obtained and is displayed in Appendix II.

First, from the first three eigenfunctions of weekdays and Sunday, we see that the major mode of variation reflected by the first eigenfunction shows that the hourly step count at the peak hours differs a lot. This simply indicates that most people are much more active than the activity pattern reflected by the mean function. The second eigenfunction shows that some users are less active at one peak while much more active at the other peak, suggesting that some users' step-count function may only have one peak or a more dominant peak. The third eigenfunction is the most interesting one. It suggests that some user are less active at the two peaks and they are also more active at some time in between the peaks. The third eigenfunction indicates that there are some users that may have step-count function with three peaks. Notice that this matches the result of Fisher-Rao warping method except that the Fisher-Rao warping method believes that the three-peak step-count function is the latent function.

Things are different for Saturday activity data, landmark registration and pairwise warping believes that the latent function of users' step-count on Saturday has only one peak and the first eigenfunction shows that the major mode of variation, similar to the weekdays, exists during the peak activity hours. The second eigenfunction reveals that, instead of being quite active throughout the whole Saturday, some user tends to be more active in Saturday morning and less so in the afternoon but there is still only one peak active hour on Satur-

day. The third eigenfunction shows that there are also some people who have a two-peak step-count function even on Saturdays and this is True after visual inspection of the warped datasets. Some typical user step-count function on Saturday looks like

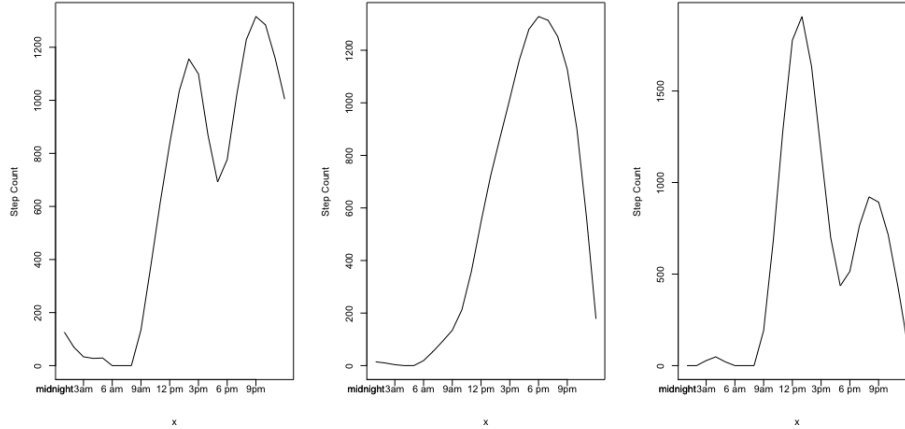


Figure 12: Typical Saturday Step-Count Functions. The right function is the standard two-peak function. The middle function is the typical one-peak function. The left function is a two-peak function with the first peak being dominant over the second one.

These typical step-count functions further proves that the conclusion we drawn from the three leading eigenfunctions are true and these three plot will later be compared with the typical Saturday step-count functions obtained from Fisher-Rao warping.

3.6.2 Fisher-Rao Warping

The warped functions displayed in Figure 4 and Figure 6 shows that the Fisher-Rao method, unlike the other two warping methods, believe that the latent function of weekdays and Sunday generally have three peaks, which matches the information conveyed by the third eigenfunction from the landmark and pairwise warping datasets. Despite the difference in the latent function, the information contained in the eigenfunctions of the Fisher-Rao warped dataset is quite similar. The first eigenfunction shows that the main mode of variation exists in the peak hours. The second eigenfunction shows that some user are more active at one specific peaks while less active at the other two or it can also suggest that some user may have

a two-peak step-count function as suggested by the other two warping methods. The third eigenfunction indicates that there are also some user who has a single-peak step-function. Therefore, the major difference between the Fisher-Rao method and the other two warping methods is that they do not agree on the main mode of variation. Other than that, they contain very similar information.

The Saturday step-count function given by the Fisher-Rao method is quite noisy. There are several peaks albeit the general trend is a single-peak function. Despite the noise, however, the Saturday step-count mean function and the leading eigenfunctions of the Fisher-Rao method conveys almost the same information as that of the other two methods. Similar to the results of the other two methods, the first eigenfunction shows that, as the other days, the largest variation occurs at the peak. The second eigenfunction indicates that some user's, although have one-peak step-count function, the activity peak occurs in the morning instead of in the middle of the day. The third eigenfunction shows that some user also have a two-peak step-count function on Saturdays. Notice that the eigenfunction of Saturday data is identical to that of the other two warping methods. The typical functions of users' Saturday step-count data looks like:

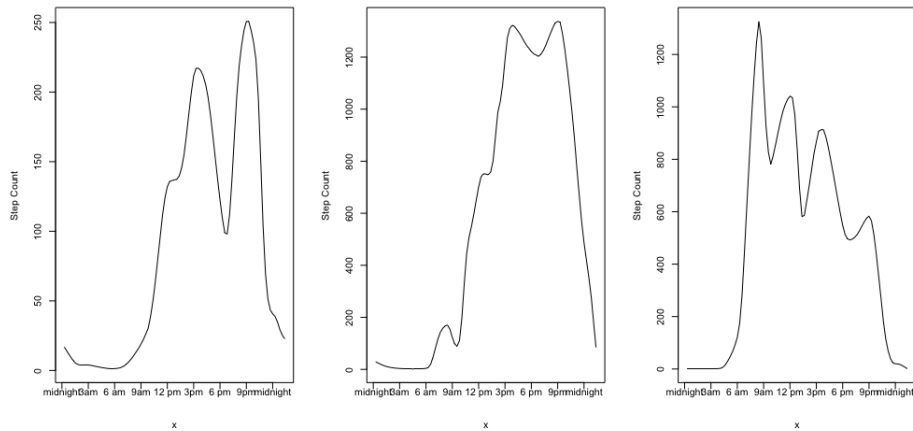


Figure 13: Typical Saturday Step-Count Functions. The right function is the standard two-peak function. The middle function is the typical one-peak function. The left function is a two-peak function with the first peak being dominant over the second one.

Notice that, if we compare the typical step-count function of user activity using Fisher-Rao warping method with the previous plot of the other two methods, we can see that the typical pattern is indeed very similar. This further indicates that all three methods conveys very similar information regarding user' step count data. Yet, they differ in determining the latent function. The landmark registration and pairwise warping method believe that the latent function is a two-peak function while the Fisher-Rao method believes that the true latent function is a three-peak function.

References

- [1] SicientDirect. *Wearable Technology*. Available at <https://www.sciencedirect.com/topics/engineering/wearable-technology>.
- [2] Alicia Phaneuf. Latest trends in medical monitoring devices and wearable health technology, Jan 2020.
- [3] Marc Bächlin, M Plotnik, D Roggen, N Giladi, JM Hausdorff, and G Tröster. A wearable system to assist walking of parkinson s disease patients. *Methods of information in medicine*, 49(01):88–95, 2010.
- [4] F. Yang and L. Zhang. Real-time human activity classification by accelerometer embedded wearable devices. In *2017 4th International Conference on Systems and Informatics (ICSAI)*, pages 469–473, 2017.
- [5] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer New York, 1997.
- [6] P. Kokoszka and M. Reimherr. *Introduction to Functional Data Analysis*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 2017.
- [7] M. Loève. Fonctions aléatoires à décomposition orthogonale exponentielle. 84:159–162, 05 2020.
- [8] K. Karhunen. *Zur Spektraltheorie stochastischer Prozesse*. Annales Academiae scientiarum Fennicae. Series A. 1, Mathematica-physica. 1946.
- [9] A Kneip, Xiaochun Li, KB MacGibbon, and JO Ramsay. Curve registration by local regression. *Canadian Journal of Statistics*, 28(1):19–29, 2000.
- [10] Anuj Srivastava, Wei Wu, Sebastian Kurtek, Eric Klassen, and J. S. Marron. Registration of Functional Data Using Fisher-Rao Metric. 2011.

- [11] Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Mueller. Review of functional data analysis, 2015.
- [12] J. S. Marron, James O. Ramsay, Laura M. Sangalli, and Anuj Srivastava. Functional data analysis of amplitude and phase variation. *Statistical Science*, 30(4):468–484, 2015.
- [13] Theo Gasser and Alois Kneip. Searching for structure in curve samples. *Journal of the American Statistical Association*, 90(432):1179–1188, 1995.
- [14] Alois Kneip and Theo Gasser. Statistical Tools to Analyze Data Representing a Sample of Curves Author (s): Alois Kneip and Theo Gasser Source : The Annals of Statistics , Vol . 20 , No . 3 (Sep . , 1992), pp . 1266-1305 Published by : Institute of Mathematical Statistics Stable UR. *The Annals of Statistics*, 20(3):1266–1305, 1992.
- [15] Kongming Wang and Theo Gasser. Alignment of curves by dynamic time warping. *Annals of Statistics*, 25(3):1251–1276, 1997.
- [16] Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1):43–49, 1978.
- [17] Theo Wang, Kongming and Gasser. SYNCHRONIZING SAMPLE CURVES NON-PARAMETRICALLY. *The Annals of Statistics*, 27(2):439–460, 1999.
- [18] J. O. Ramsay and Xiaochun Li. Curve registration. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 60(2):351–363, 1998.
- [19] Gareth M. James. Curve alignment by moments. *The Annals of Applied Statistics*, 1(2):480–501, 2007.
- [20] C. Radhakrishna Rao. Information and the Accuracy Attainable in the Estimation of Statistical Parameters. 1992.

- [21] Xueli Liu and Hans Georg Müller. Functional convex averaging and synchronization for time-warped random curves. *Journal of the American Statistical Association*, 99(467):687–699, 2004.
- [22] Simone Vantini. On the definition of phase and amplitude variability in functional data analysis. *Test*, 21(4):676–696, 2012.
- [23] Jason Cleveland, Weilong Zhao, and Wei Wu. Robust template estimation for functional data with phase variability using band depth. *Computational Statistics and Data Analysis*, 2018.
- [24] Sebastian Kurtek, Anuj Srivastava, and Wei Wu. Signal estimation under random time-warpings and nonlinear signal alignment. *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011*, 32306(1):1–9, 2011.
- [25] Hongjun Choi, Qiao Wang, Meynard Toledo, Pavan Turaga, Matthew Buman, and Anuj Srivastava. Temporal alignment improves feature quality: An experiment on activity recognition with accelerometer data. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2018-June:462–470, 2018.
- [26] Jason Cleveland, Wei Wu, and Anuj Srivastava. Norm-preserving constraint in the Fisher–Rao registration and its application in signal estimation. *Journal of Nonparametric Statistics*, 28(2):338–359, apr 2016.
- [27] J. Derek Tucker, Wei Wu, and Anuj Srivastava. Generative models for functional data using phase and amplitude separation. *Computational Statistics and Data Analysis*, 61:50–66, 2013.
- [28] Rong Tang, Hans Georg Muller, and H Müller. Pairwise curve synchronization for high-dimensional data., *Biometrika* 95, 875–889. *Mathematical Reviews (MathSciNet): MR2461217 Digital Object Identifier: doi*, 10(4):875–889, 2008.

- [29] Cecilia Earls and Giles Hooker. Variational bayes for functional data registration, smoothing, and prediction. *Bayesian Analysis*, 12(2):557–582, 2017.
- [30] Julia Wrobel, Vadim Zipunnikov, Jennifer Schrack, and Jeff Goldsmith. Registration for exponential family functional data. *Biometrics*, 75(1):48–57, 2019.
- [31] Daniel Gervini and Theo Gasser. Self-modelling warping functions. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 66(4):959–971, 2004.
- [32] Anirvan Chakraborty and Victor M. Panaretos. Functional Registration and Local Variations: Identifiability, Rank, and Tuning. 2017.
- [33] Alois Kneip and James O. Ramsay. Combining registration and fitting for functional models. *Journal of the American Statistical Association*, 103(483):1155–1165, 2008.

Appendix I. Warping Plots

Landmark Registration

1-Step Landmark Registration

To analyze the result of 1-step landmark registration, we first observe the resulting aligned step-count functions and their corresponding warping functions.

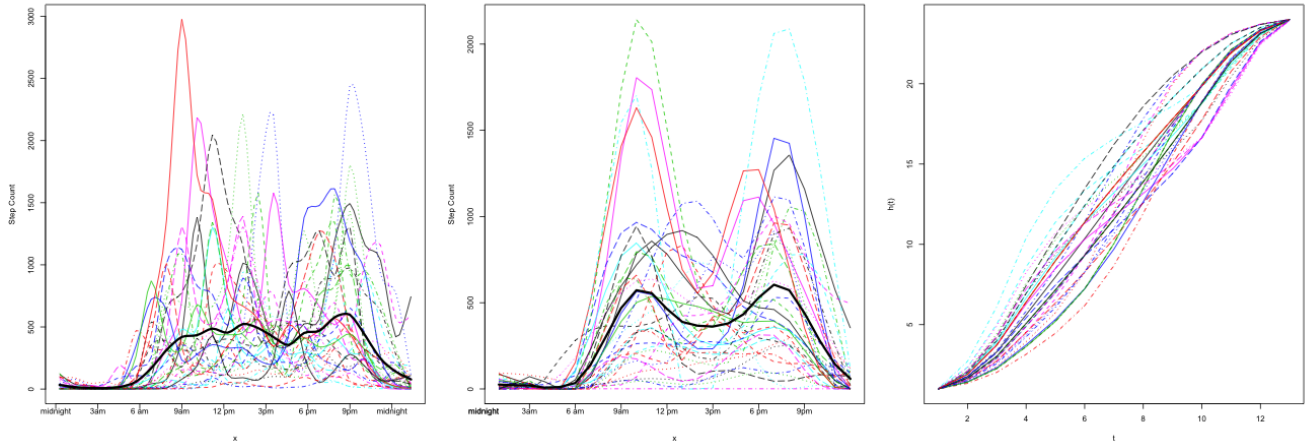


Figure 14: 1-Step Landmark Registration Result for Step-Count Data on Monday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

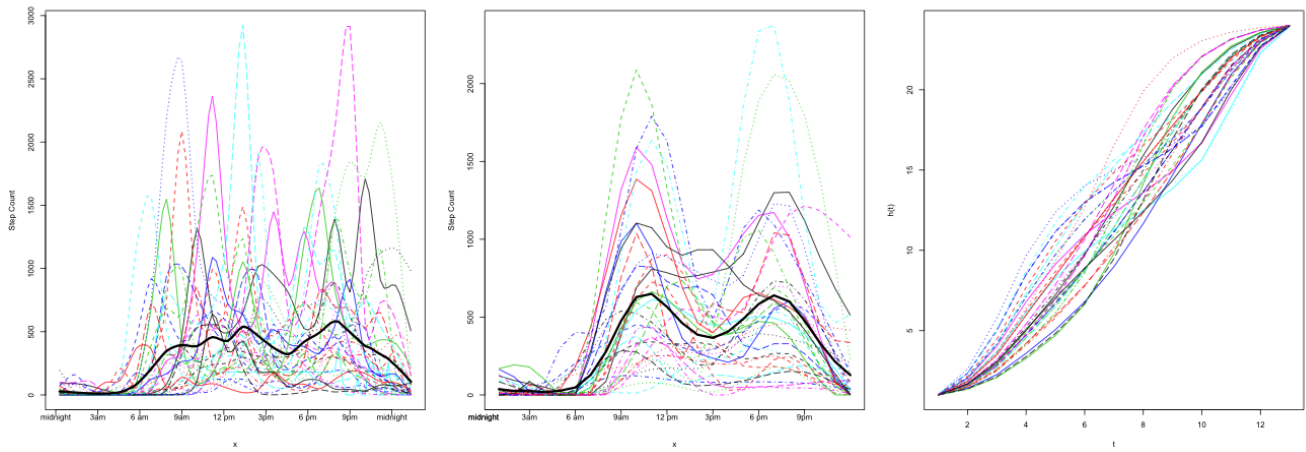


Figure 15: 1-Step Landmark Registration Result for Step-Count Data on Tuesday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

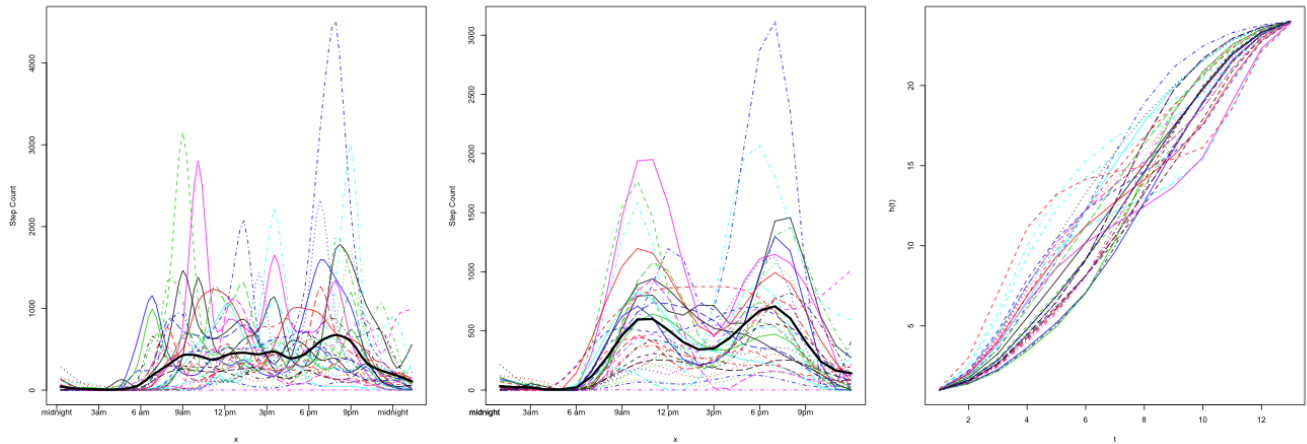


Figure 16: 1-Step Landmark Registration Result for Step-Count Data on Wednesday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

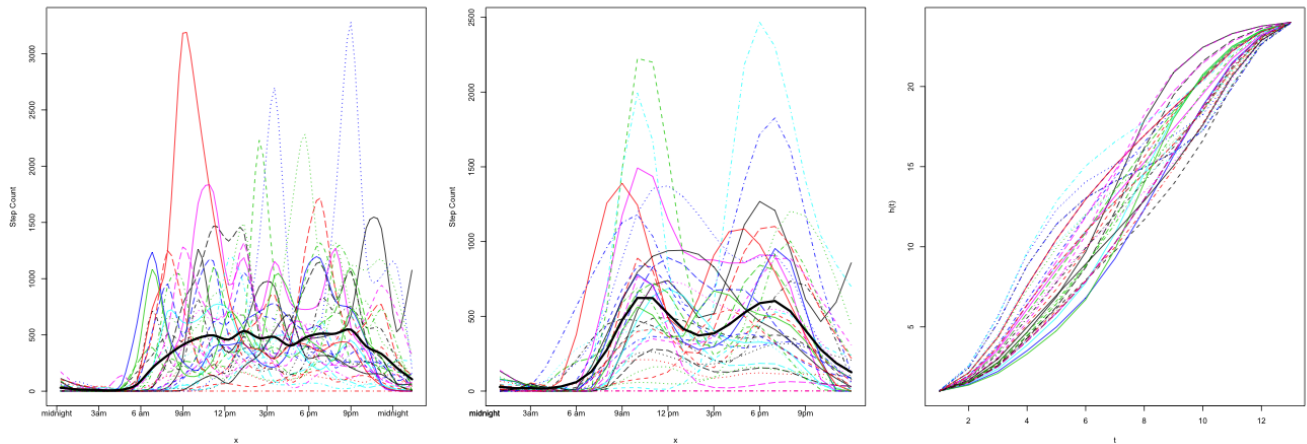


Figure 17: 1-Step Landmark Registration Result for Step-Count Data on Thursday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

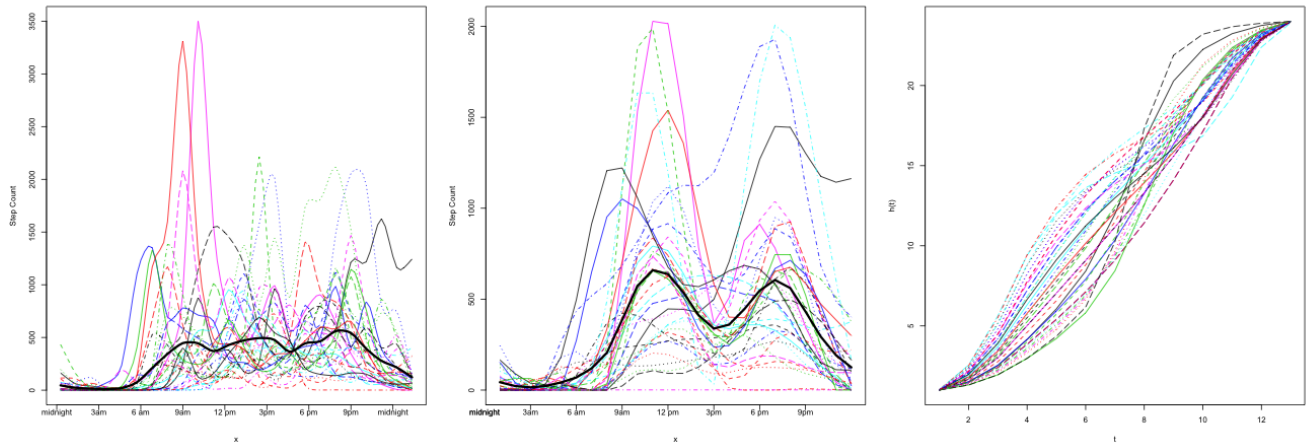


Figure 18: 1-Step Landmark Registration Result for Step-Count Data on Friday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

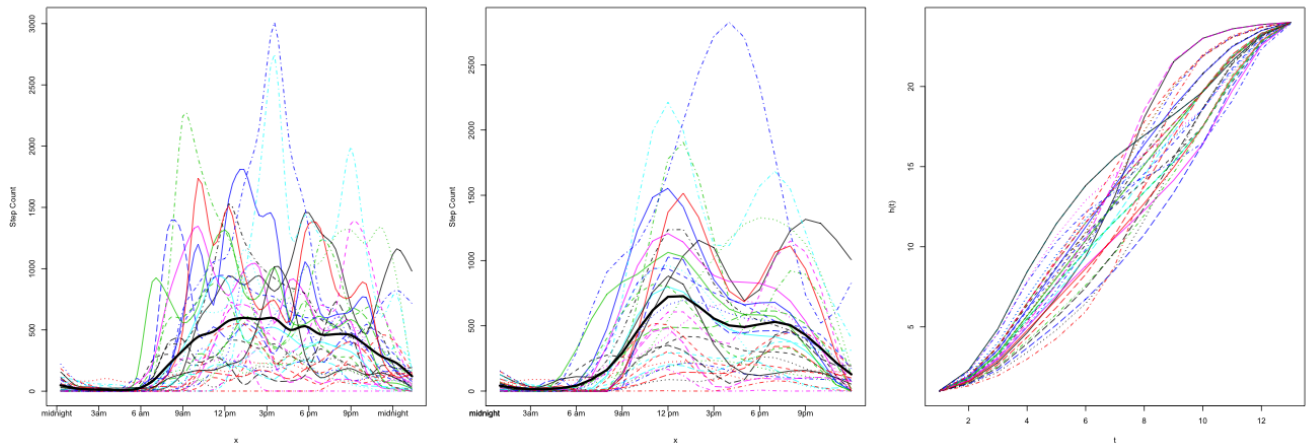


Figure 19: 1-Step Landmark Registration Result for Step-Count Data on Saturday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

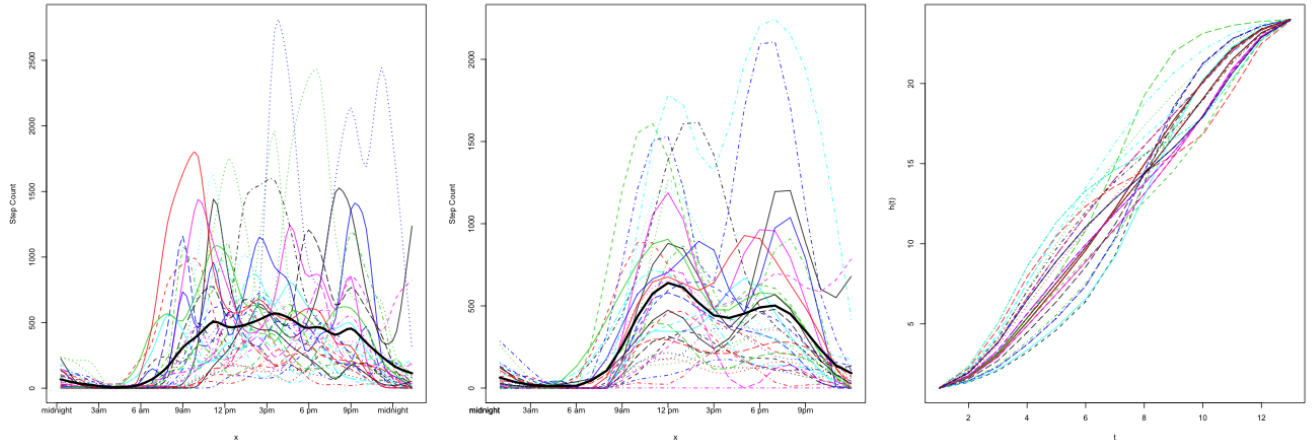


Figure 20: 1-Step Landmark Registration Result for Step-Count Data on Sunday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

2-Step Landmark Registration

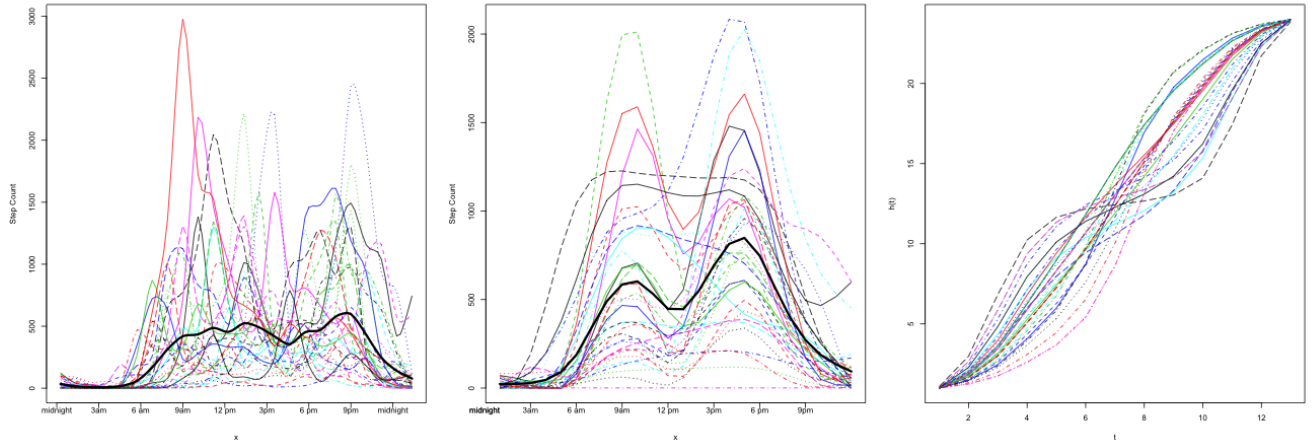


Figure 21: 2-Step Landmark Registration Result for Step-Count Data on Monday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

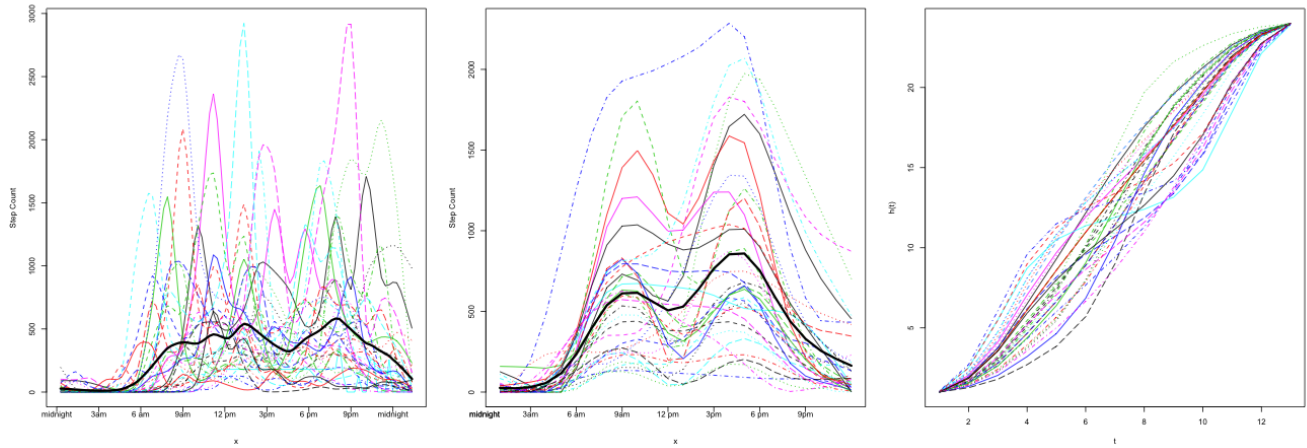


Figure 22: 2-Step Landmark Registration Result for Step-Count Data on Tuesday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

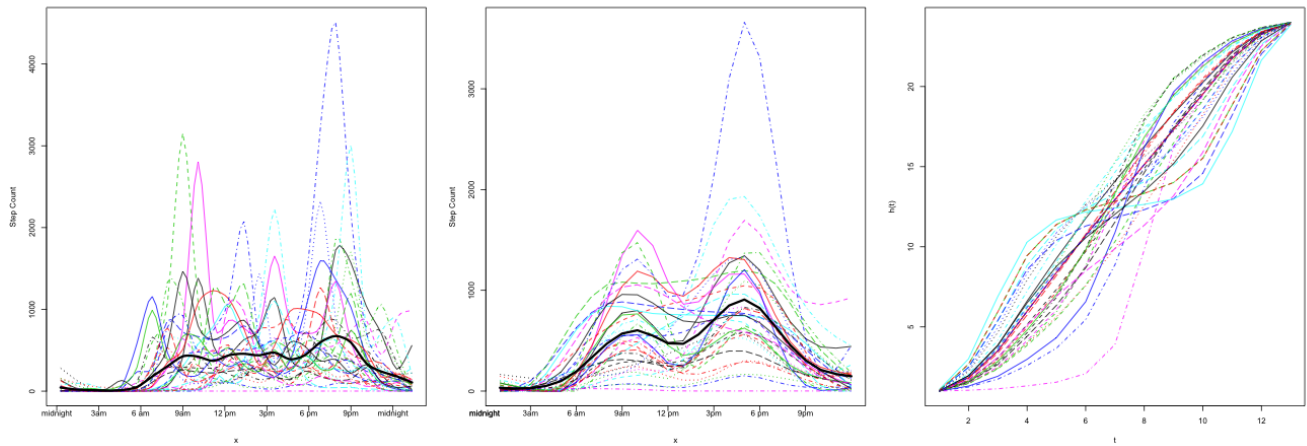


Figure 23: 2-Step Landmark Registration Result for Step-Count Data on Wednesday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

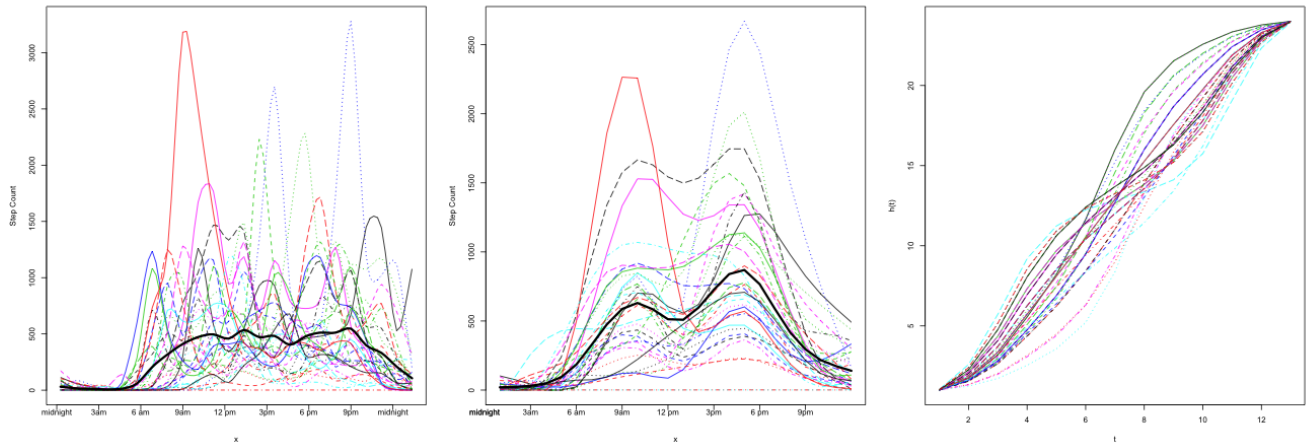


Figure 24: 2-Step Landmark Registration Result for Step-Count Data on Thursday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

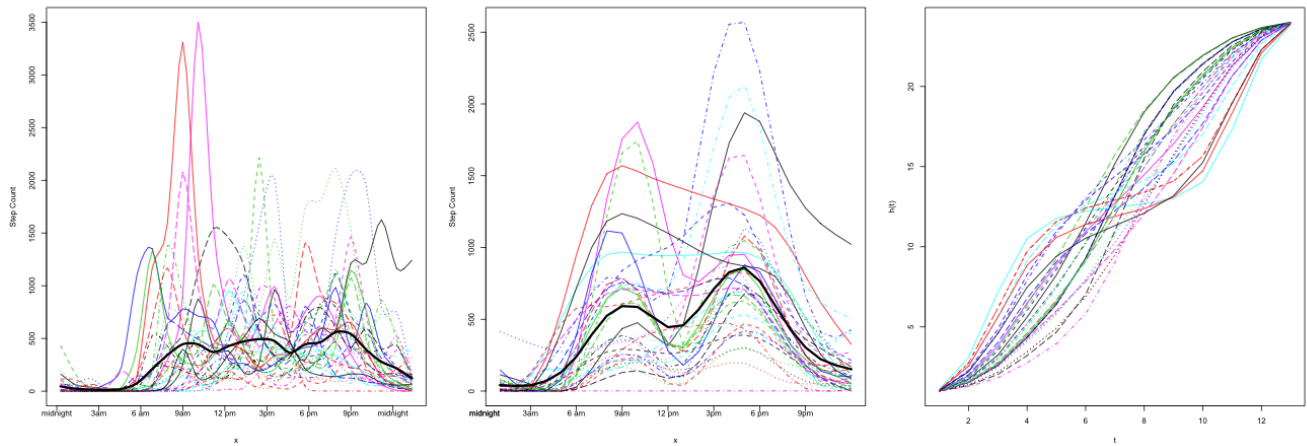


Figure 25: 2-Step Landmark Registration Result for Step-Count Data on Friday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

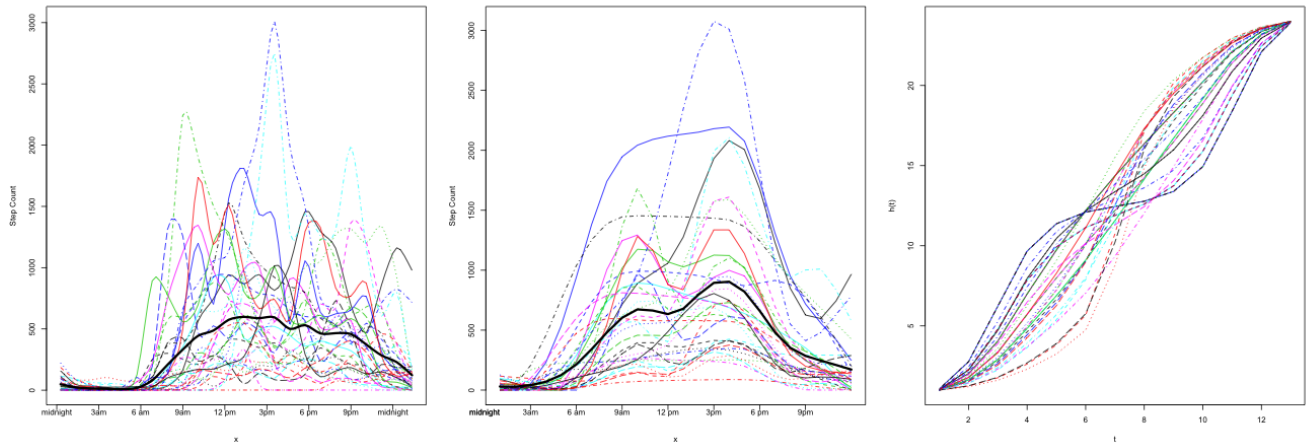


Figure 26: 2-Step Landmark Registration Result for Step-Count Data on Saturday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

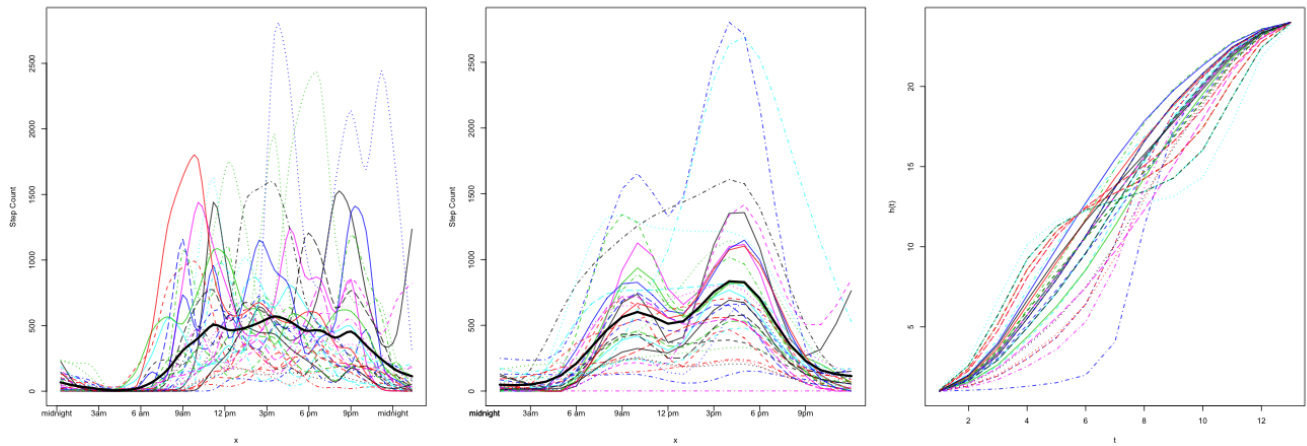


Figure 27: 2-Step Landmark Registration Result for Step-Count Data on Sunday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

Pairwise Curve Synchronization

1-Step Pairwise Curve Synchronization

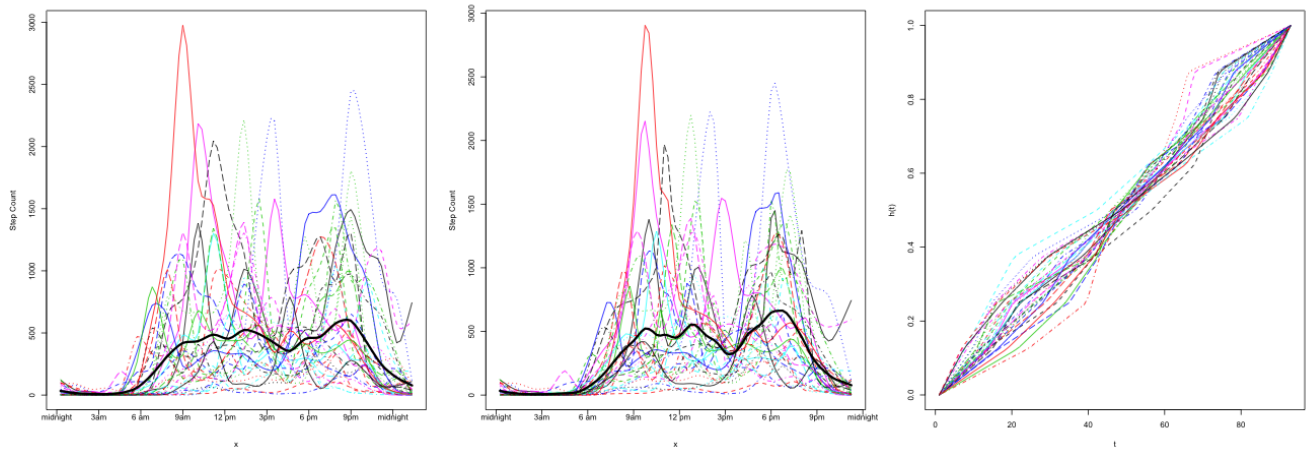


Figure 28: 1-Step Pairwise Warping Result for Step-Count Data on Monday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

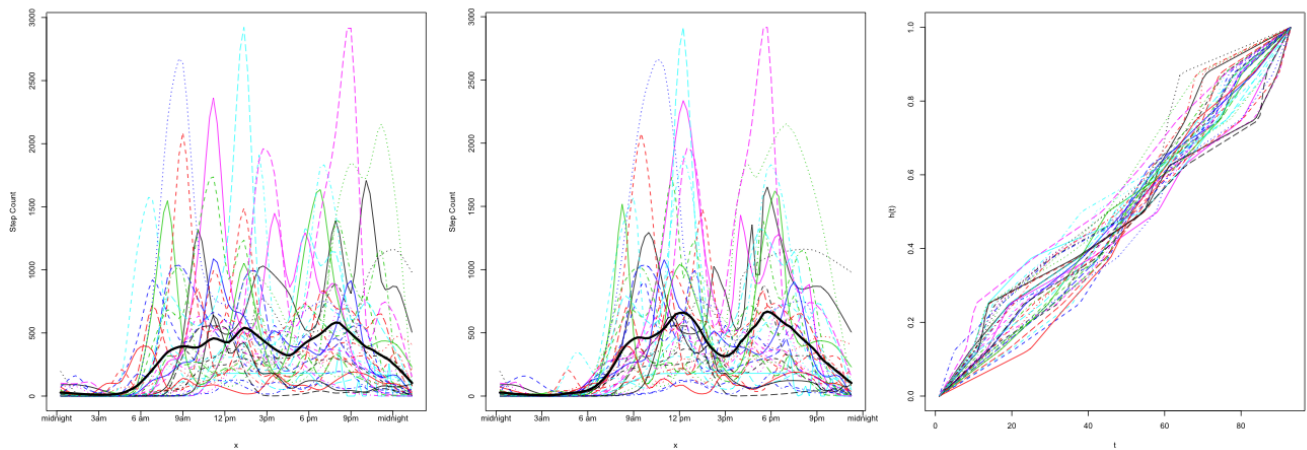


Figure 29: 1-Step Pairwise Warping Result for Step-Count Data on Tuesday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

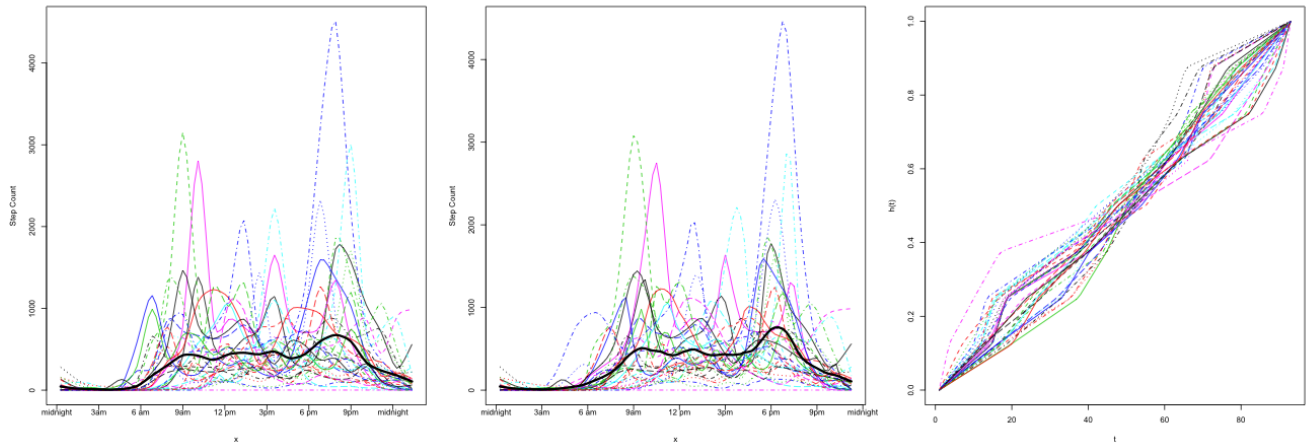


Figure 30: 1-Step Pairwise Warping Result for Step-Count Data on Wednesday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

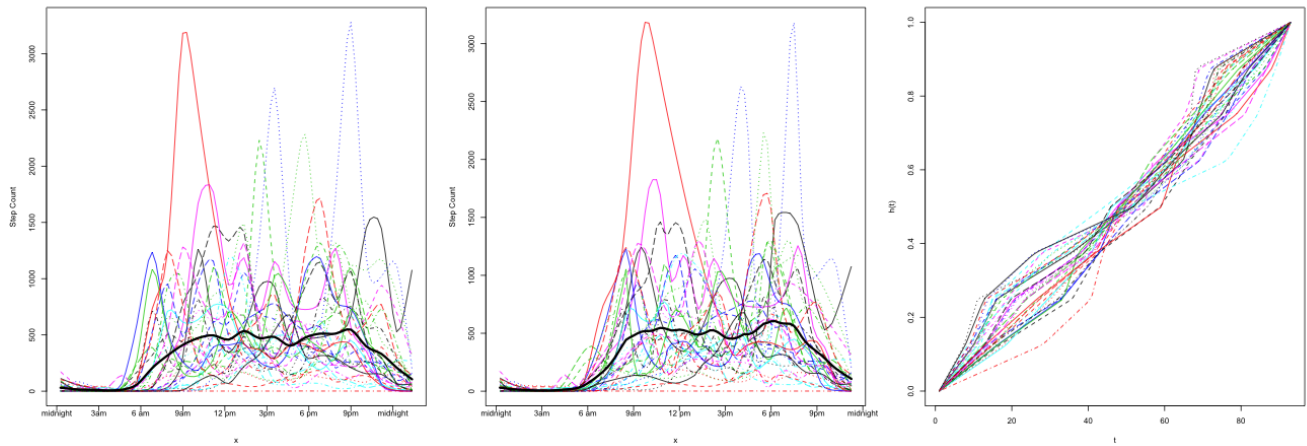


Figure 31: 1-Step Pairwise Warping Result for Step-Count Data on Thursday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

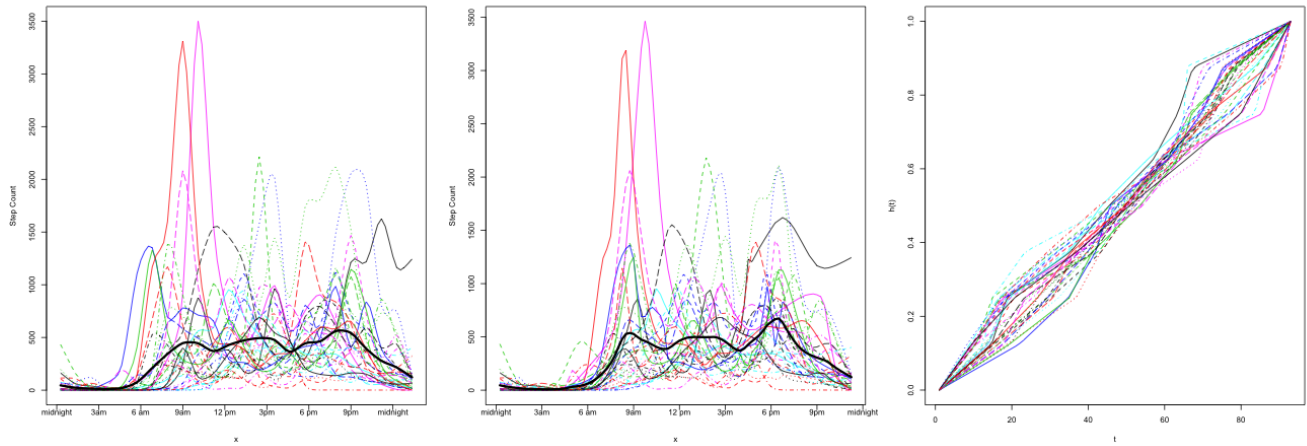


Figure 32: 1-Step Pairwise Warping Result for Step-Count Data on Friday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

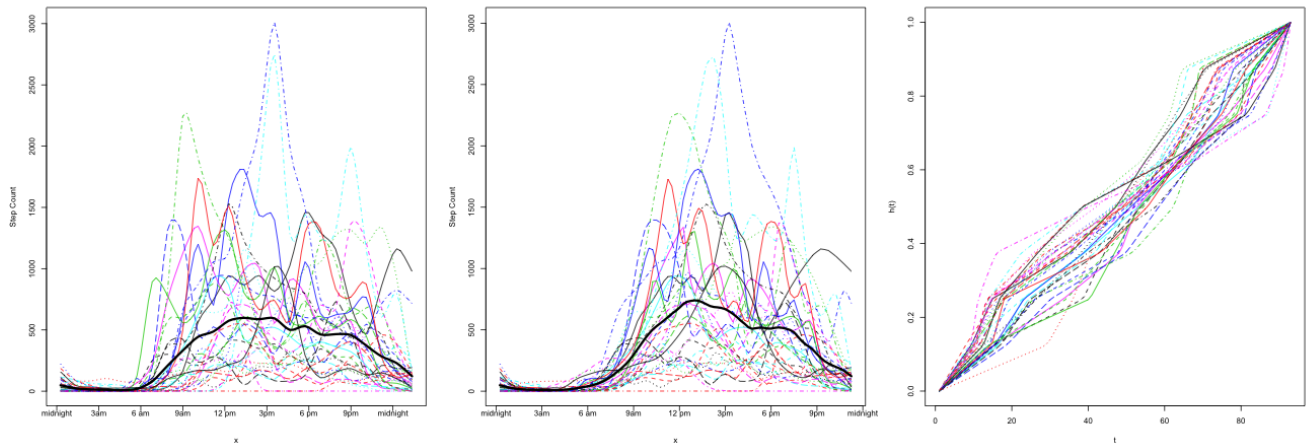


Figure 33: 1-Step Pairwise Warping Result for Step-Count Data on Saturday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

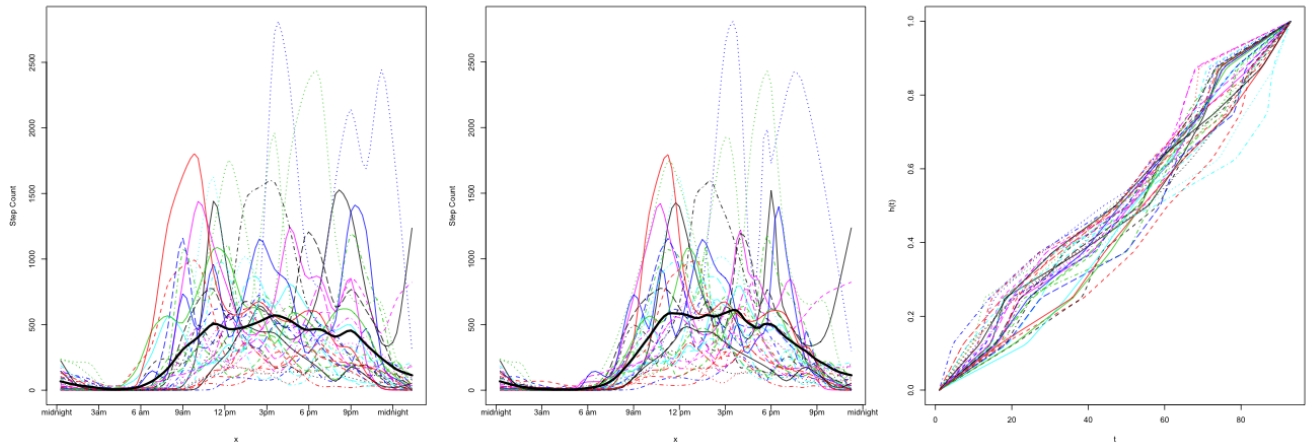


Figure 34: 1-Step Pairwise Warping Result for Step-Count Data on Sunday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

2-Step Pairwise Curve Synchronization

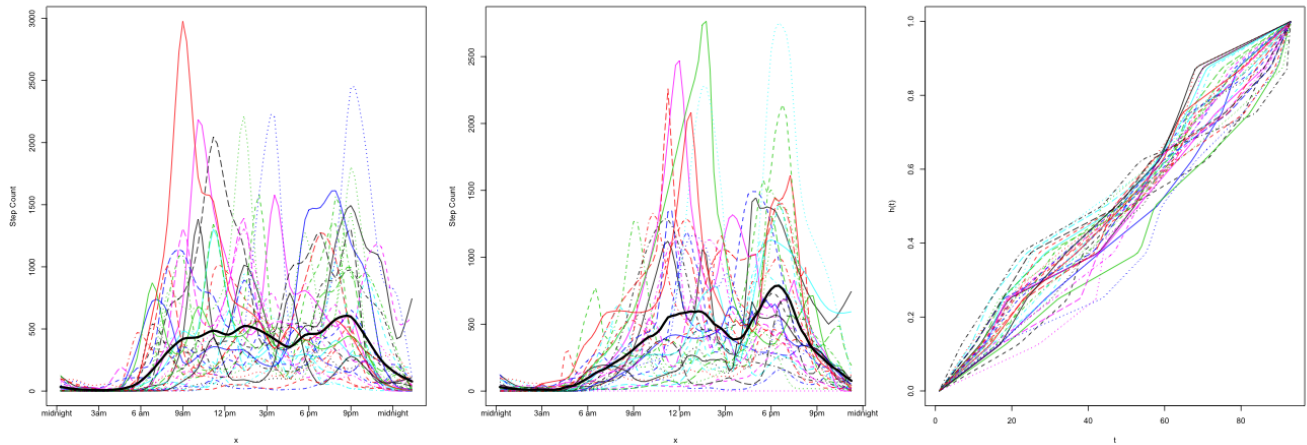


Figure 35: 2-Step Pairwise Warping Result for Step-Count Data on Monday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

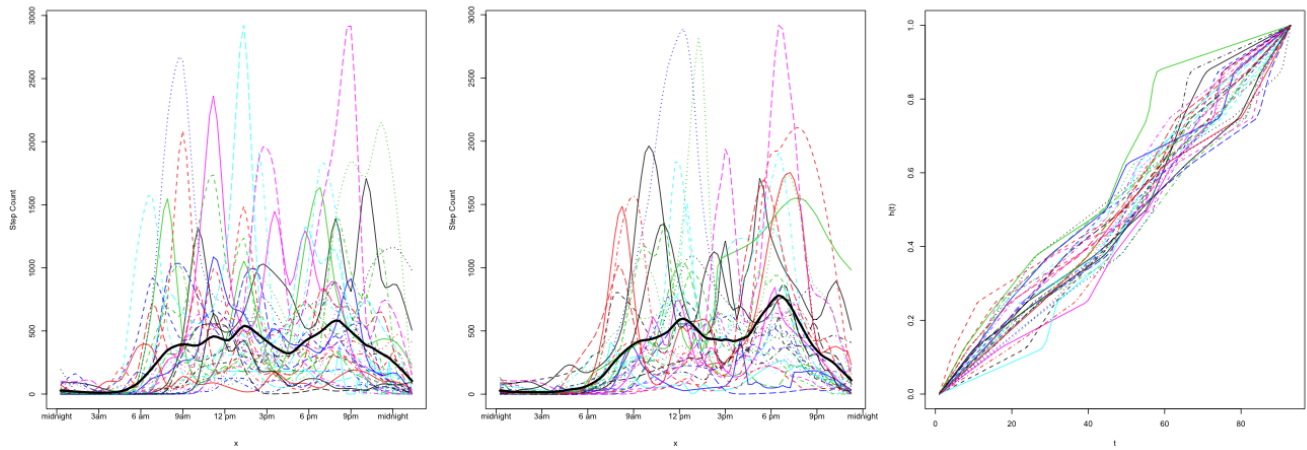


Figure 36: 2-Step Pairwise Warping Result for Step-Count Data on Tuesday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

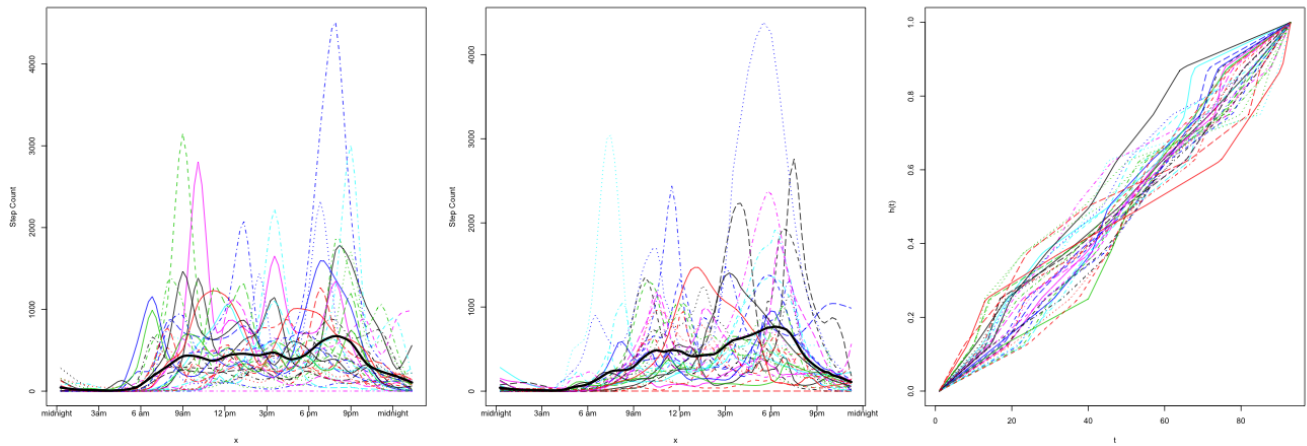


Figure 37: 2-Step Pairwise Warping Result for Step-Count Data on Wednesday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

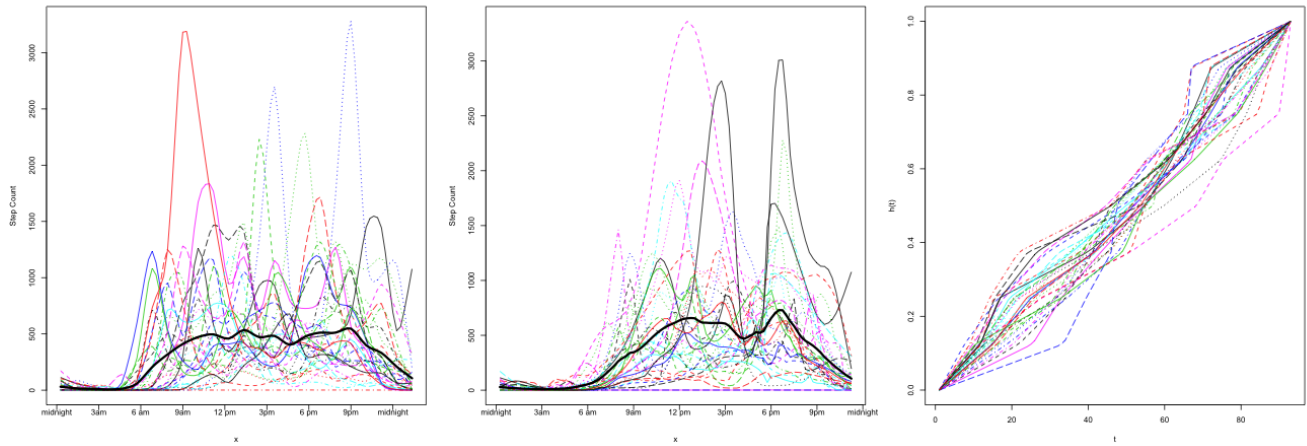


Figure 38: 2-Step Pairwise Warping Result for Step-Count Data on Thursday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

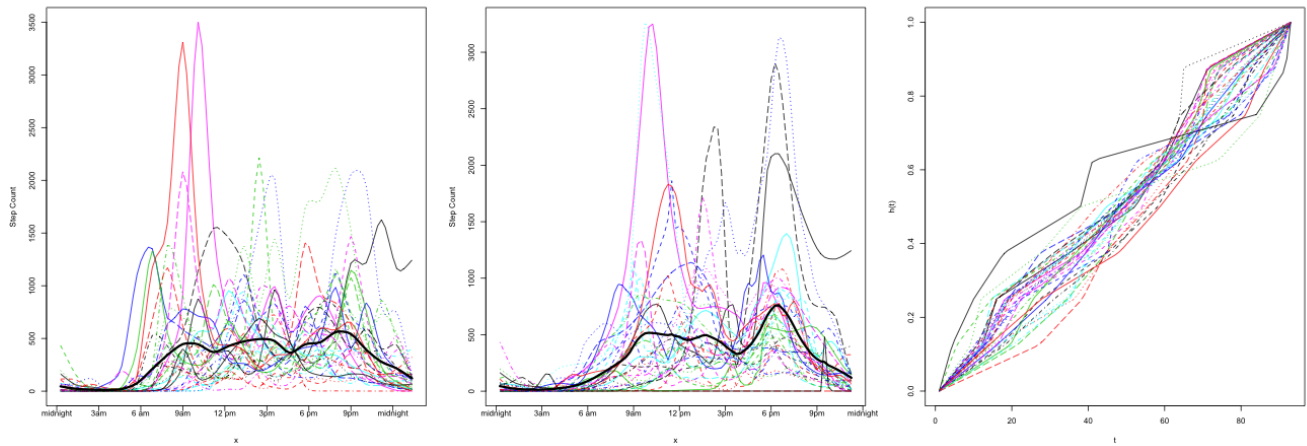


Figure 39: 2-Step Pairwise Warping Result for Step-Count Data on Friday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

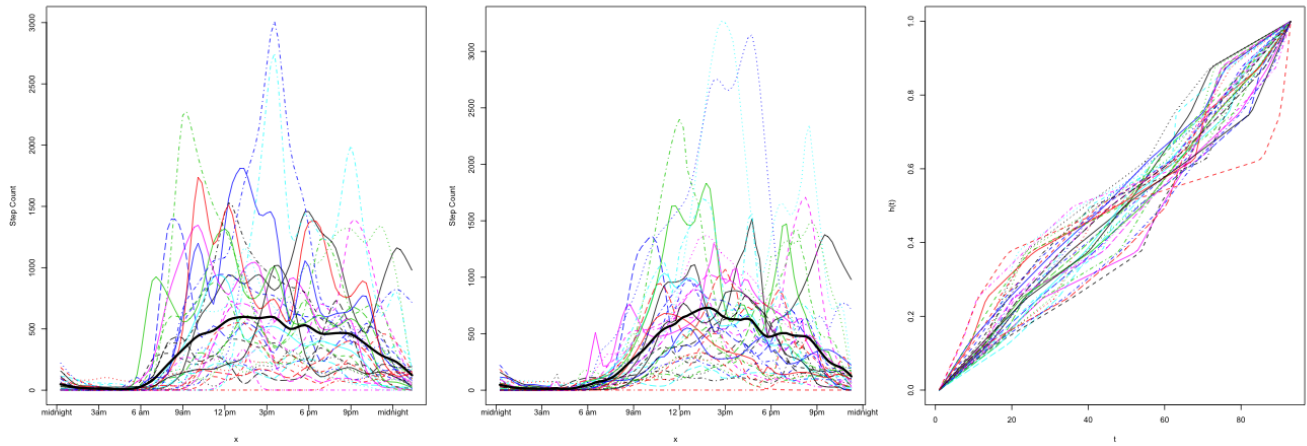


Figure 40: 2-Step Pairwise Warping Result for Step-Count Data on Saturday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

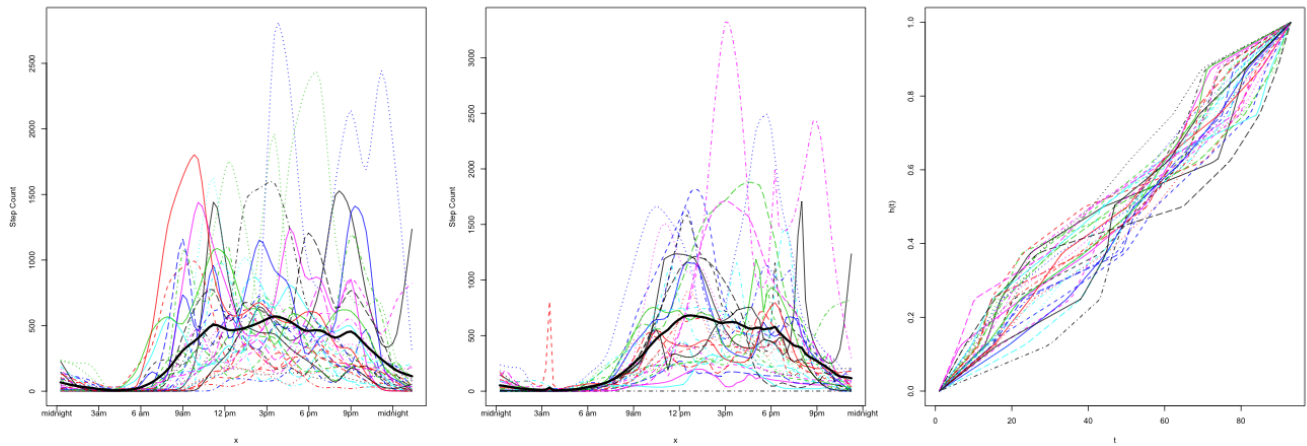


Figure 41: 2-Step Pairwise Warping Result for Step-Count Data on Sunday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

Warping with Fisher-Rao Metric

1-Step Fisher-Rao Warping

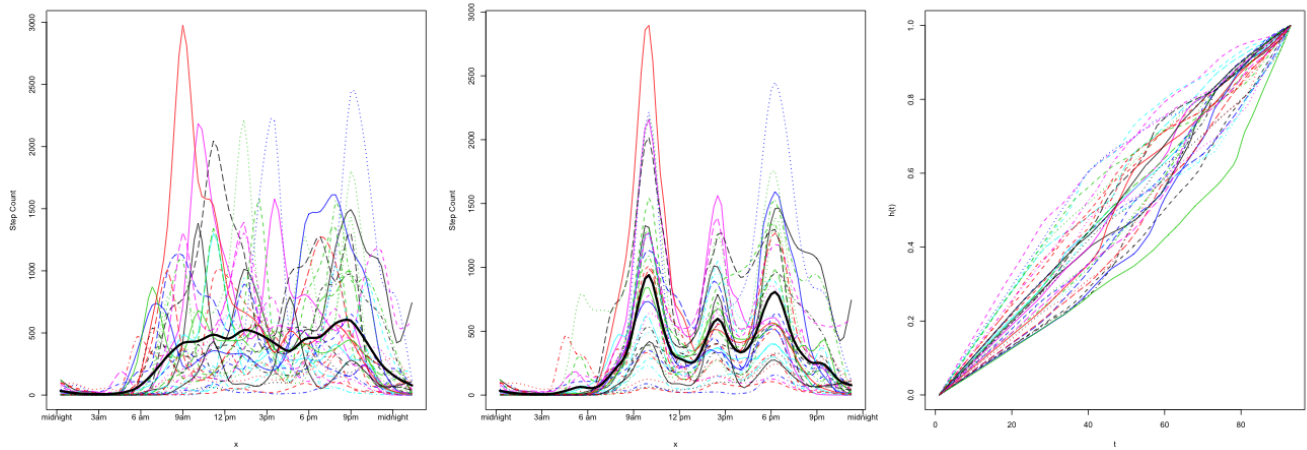


Figure 42: 1-Step Fisher-Rao Warping Result for Step-Count Data on Monday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

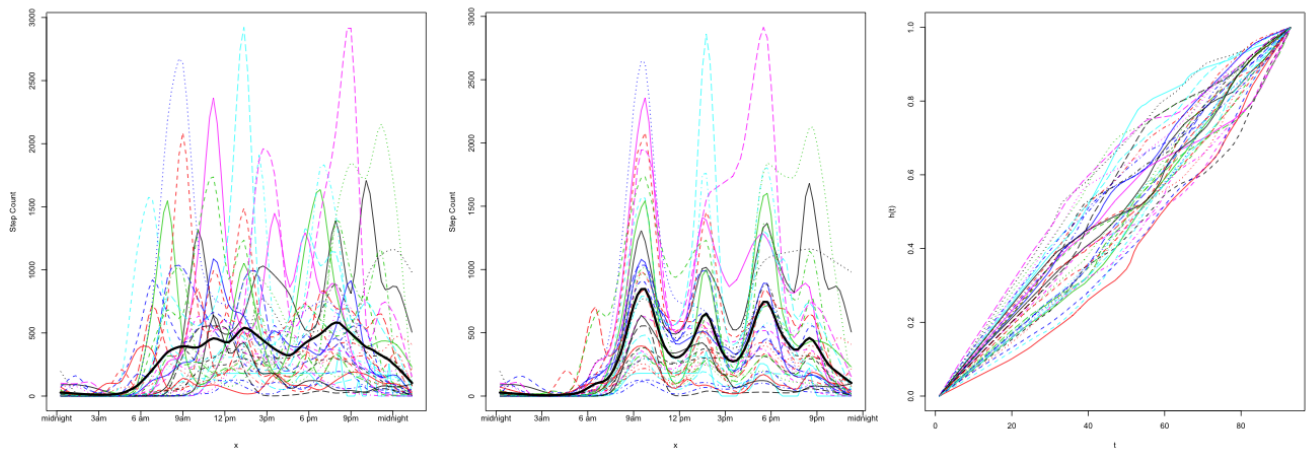


Figure 43: 1-Step Fisher-Rao Warping Result for Step-Count Data on Tuesday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

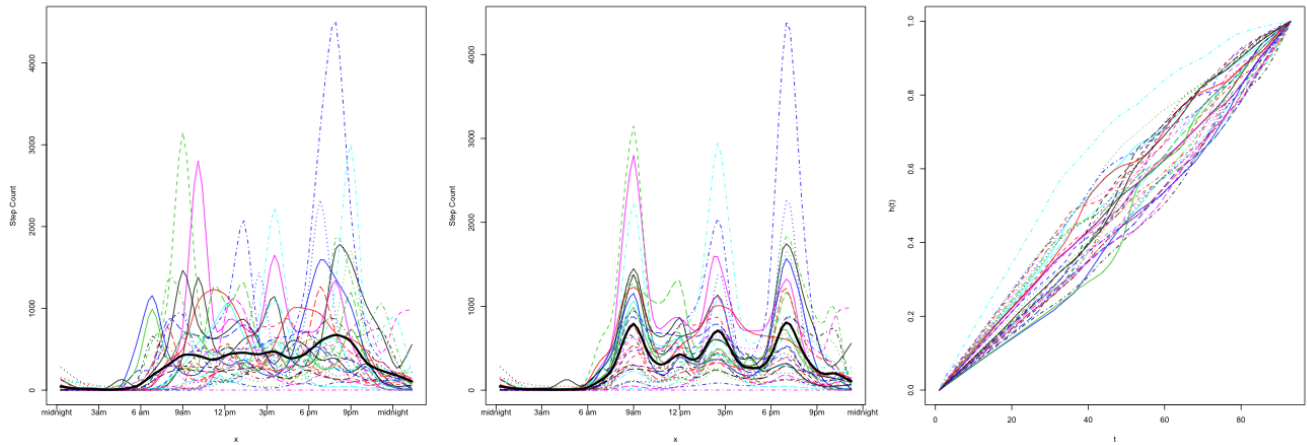


Figure 44: 1-Step Fisher-Rao Warping Result for Step-Count Data on Wednesday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

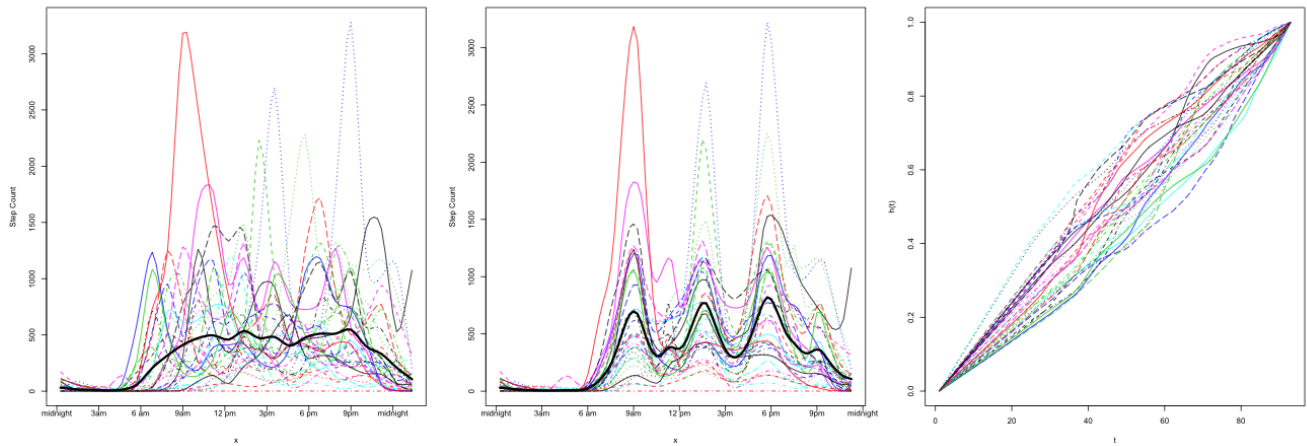


Figure 45: 1-Step Fisher-Rao Warping Result for Step-Count Data on Thursday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

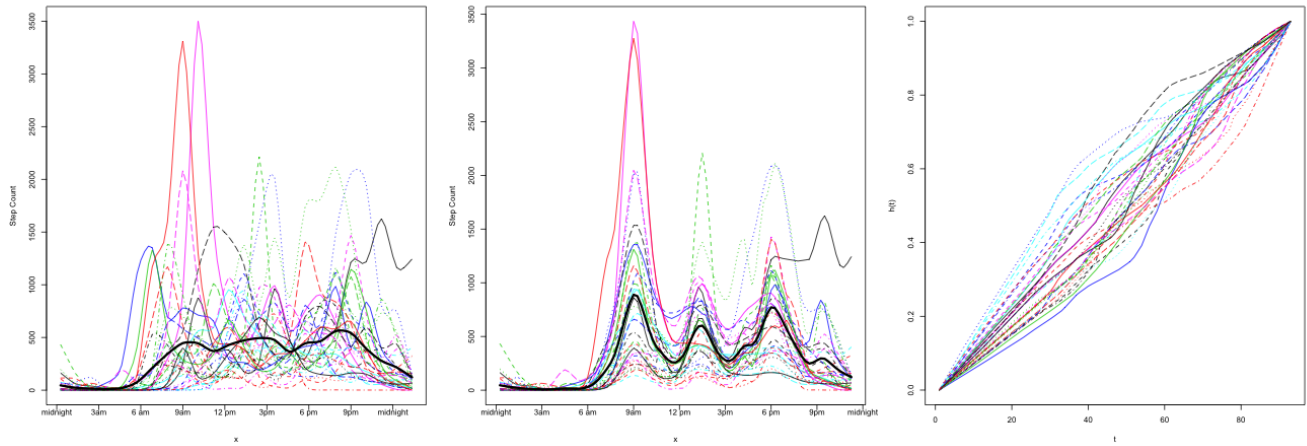


Figure 46: 1-Step Fisher-Rao Warping Result for Step-Count Data on Friday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

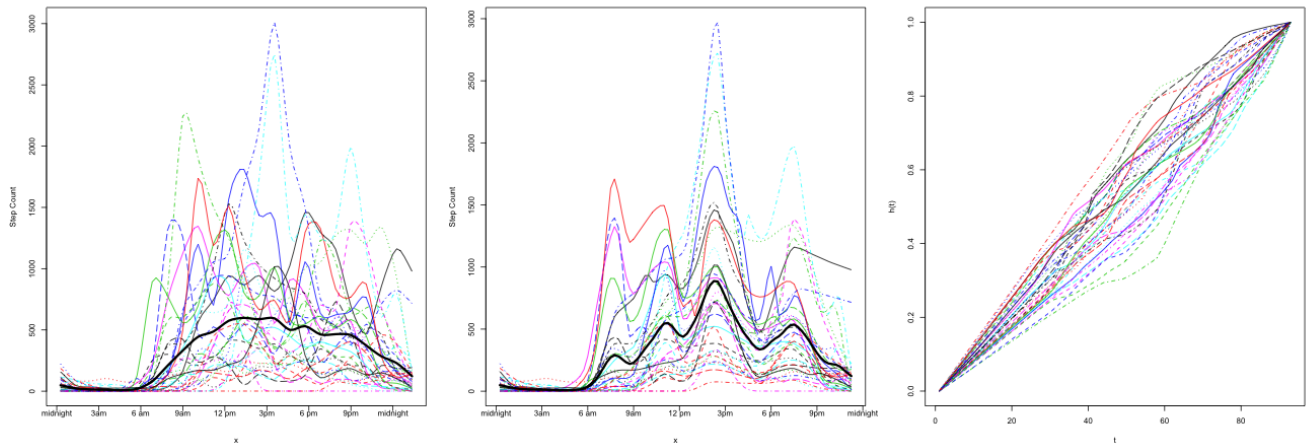


Figure 47: 1-Step Fisher-Rao Warping Result for Step-Count Data on Saturday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

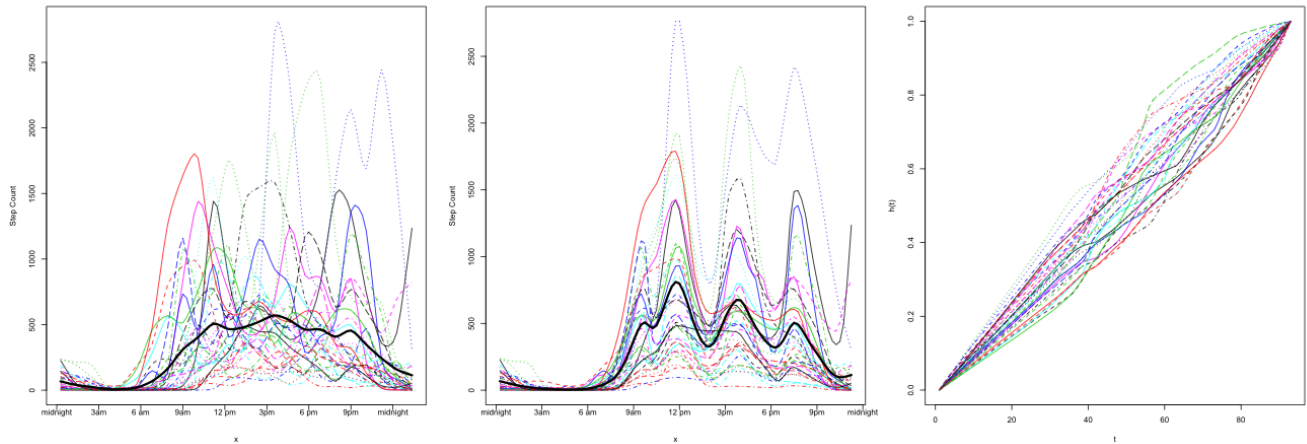


Figure 48: 1-Step Fisher-Rao Warping Result for Step-Count Data on Sunday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

2-Step Fisher-Rao Warping

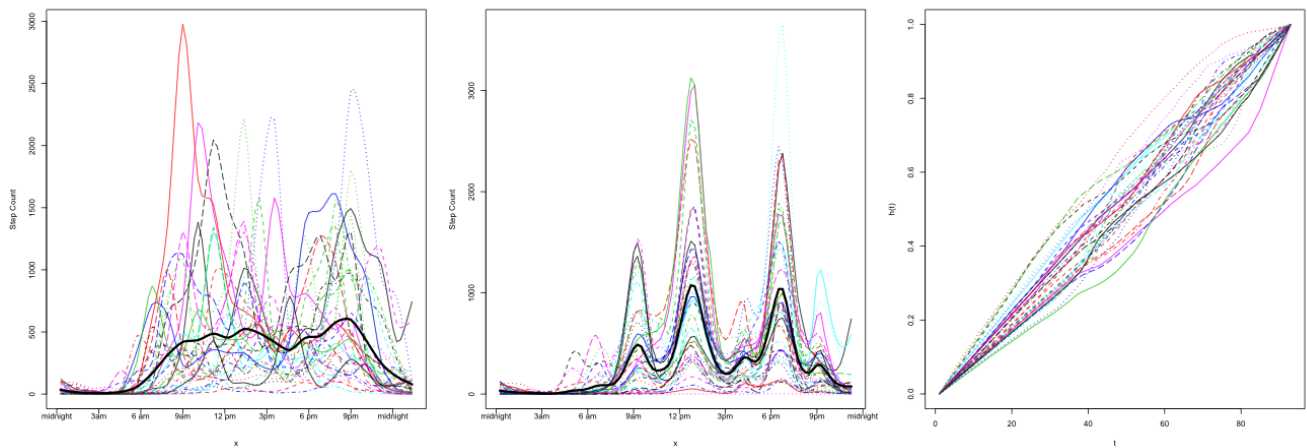


Figure 49: 2-Step Fisher-Rao Warping Result for Step-Count Data on Monday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

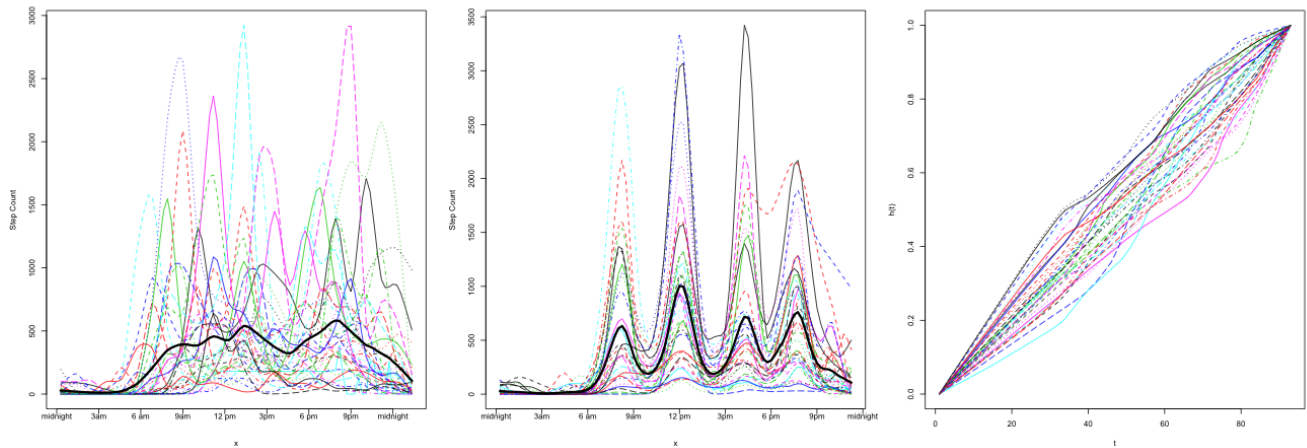


Figure 50: 2-Step Fisher-Rao Warping Result for Step-Count Data on Tuesday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

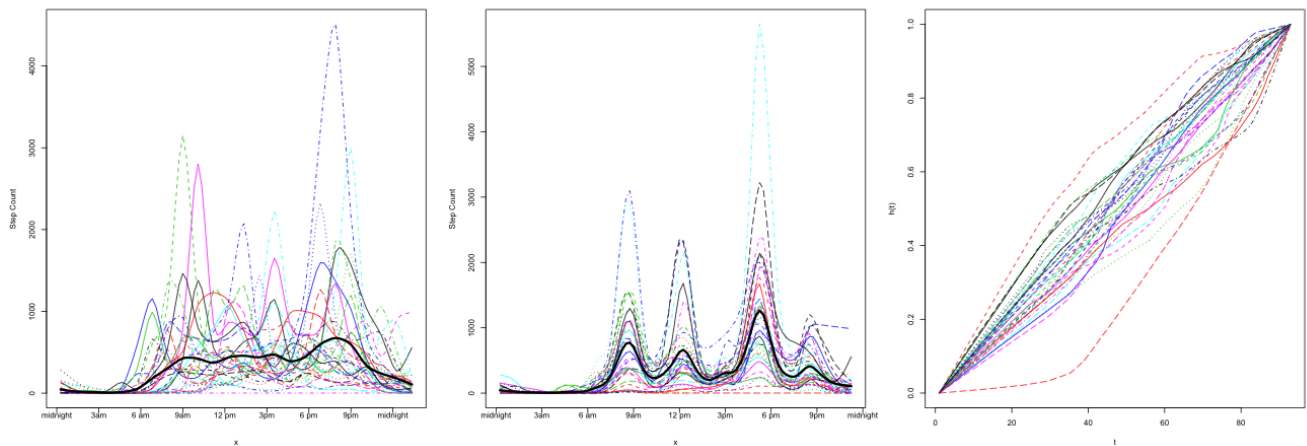


Figure 51: 2-Step Fisher-Rao Warping Result for Step-Count Data on Wednesday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

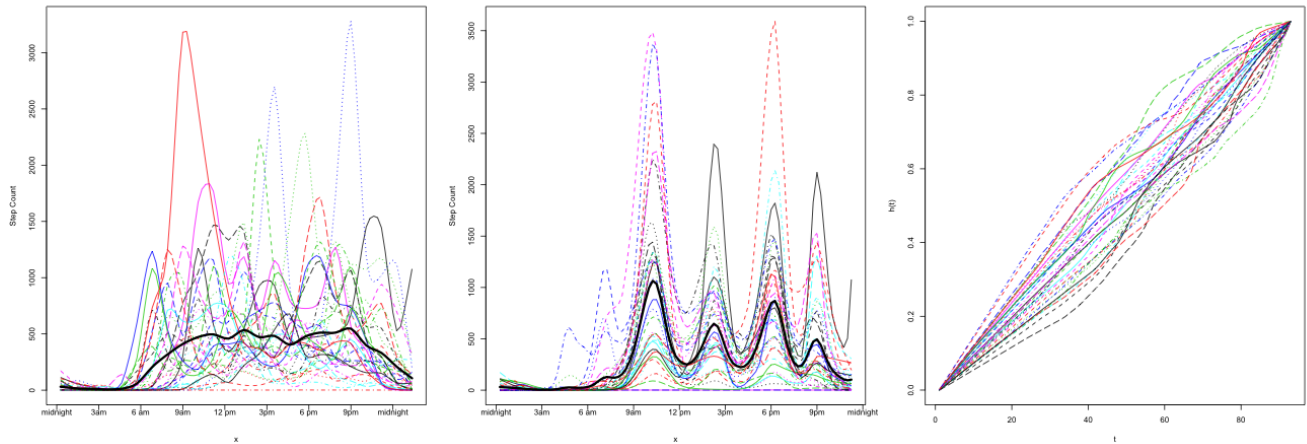


Figure 52: 2-Step Fisher-Rao Warping Result for Step-Count Data on Thursday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

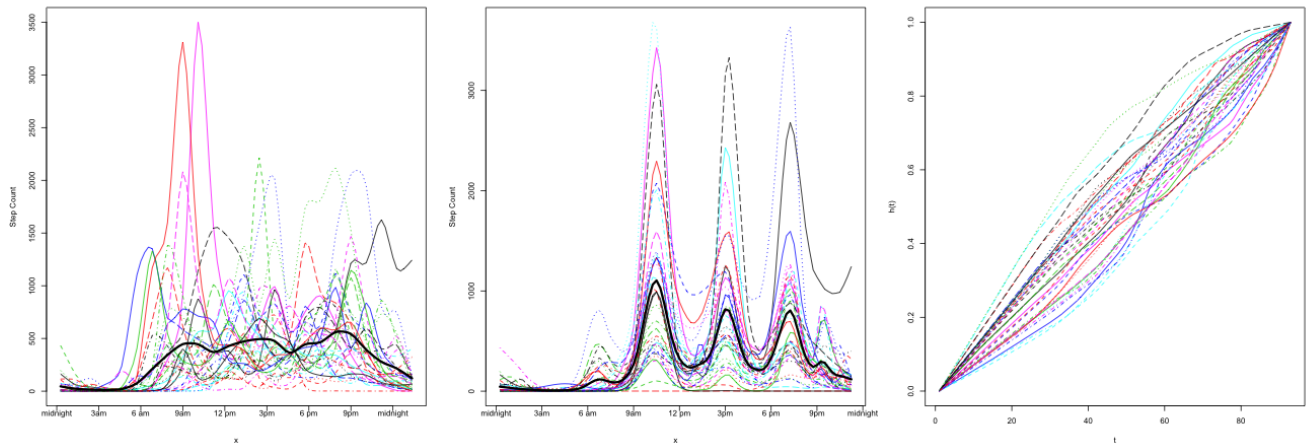


Figure 53: 2-Step Fisher-Rao Warping Result for Step-Count Data on Friday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

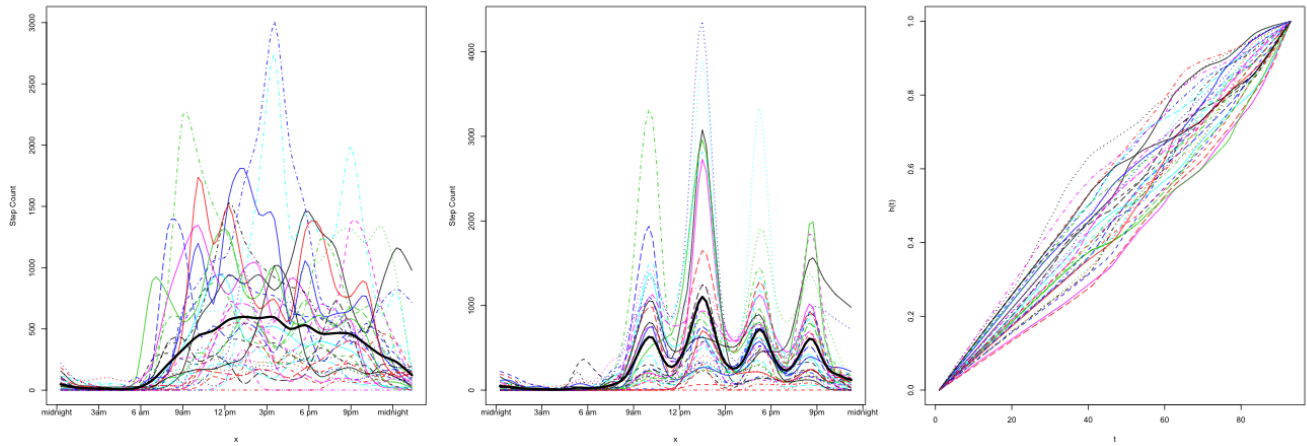


Figure 54: 2-Step Fisher-Rao Warping Result for Step-Count Data on Saturday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.

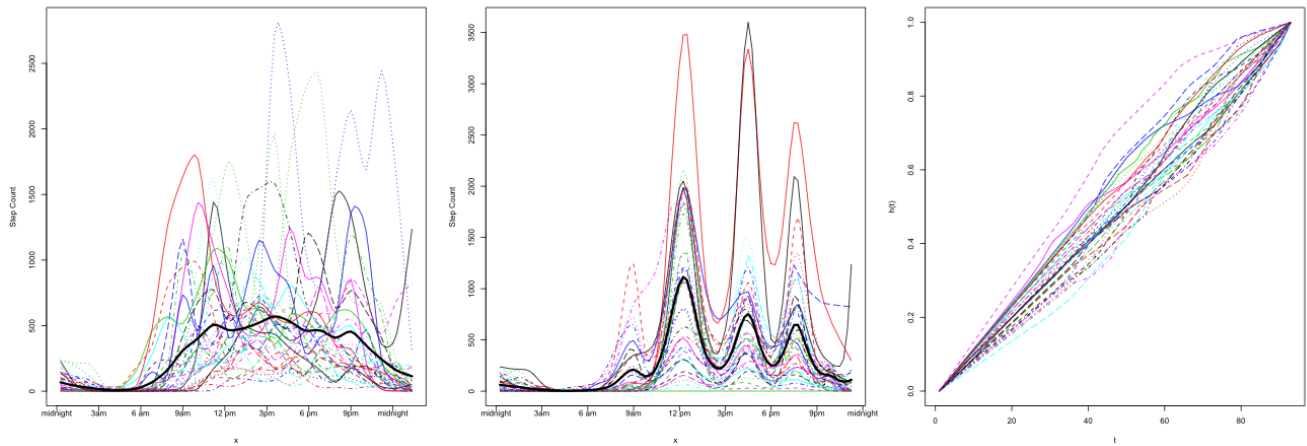


Figure 55: 2-Step Fisher-Rao Warping Result for Step-Count Data on Sunday. The thick black line represents the arithmetic mean. From right to left are original data, warped data, and corresponding warping function.