Title

A RIEMANN MANIFOLD MODEL FRAMEWORK FOR LONGITUDINAL CHANGES IN PHYSICAL ACTIVITY PATTERNS.

Journal

The Annals of Applied Statistics, 17(4)

ISSN

1932-6157

Authors

Zou, Jingjing

Lin, Tuo

Di, Chongzhi

et al.

Publication Date

2023-12-01

DOI

10.1214/23-aoas1758

Peer reviewed

HHS Public Access

Author manuscript

Ann Appl Stat. Author manuscript; available in PMC 2024 June 04.

Published in final edited form as:

Ann Appl Stat. 2023 December ; 17(4): 3216–3240. doi:10.1214/23-aoas1758.

A RIEMANN MANIFOLD MODEL FRAMEWORK FOR LONGITUDINAL CHANGES IN PHYSICAL ACTIVITY PATTERNS

Jingjing Zou[1,6], **Tuo Lin**[1], **Chongzhi Di**[2], **John Bellettiere**[1], **Marta M. Jankowska**[3], **Sheri J. Hartman**[1,6], **Dorothy D. Sears**[4,5,6], **Andrea Z. LaCroix**[1], **Cheryl L. Rock**[5], **Loki Natarajan**[1,6]

[1]Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego

[2]Division of Public Health Sciences, Fred Hutchinson Cancer Cente

[3]Department of Population Sciences, Beckman Research Institute, City of Hope

[4]College of Health Solutions, Arizona State University

[5]Department of Family Medicine, University of California, San Diego

[6]UC San Diego Moores Cancer Center

Abstract

Physical activity (PA) is significantly associated with many health outcomes. The wide usage of wearable accelerometer-based activity trackers in recent years has provided a unique opportunity for in-depth research on PA and its relations with health outcomes and interventions. Past analysis of activity tracker data relies heavily on aggregating minute-level PA records into day-level summary statistics in which important information of PA temporal/diurnal patterns is lost. In this paper we propose a novel functional data analysis approach based on Riemann manifolds for modeling PA and its longitudinal changes. We model smoothed minute-level PA of a day as one-dimensional Riemann manifolds and longitudinal changes in PA in different visits as deformations between manifolds. The variability in changes of PA among a cohort of subjects is characterized via variability in the deformation. Functional principal component analysis is further adopted to model the deformations, and PC scores are used as a proxy in modeling the relation between changes in PA and health outcomes and/or interventions. We conduct comprehensive analyses on data from two clinical trials: Reach for Health (RfH) and Metabolism, Exercise and Nutrition at

Corresponding author: Jingjing Zou (j2zou@health.ucsd.edu). 3855 Health Sciences Drive Ste. 3035, La Jolla, CA 92093-0901.

SUPPLEMENTARY MATERIAL

Supplement to "A Riemann manifold model framework for longitudinal changes in physical activity patterns" (DOI: 10.1214/23-AOAS1758SUPP; .zip).

Title: Code and additional results.

Additional Results: Document "supp_material.pdf" contains additional results of the simulation studies and data analysis.

Additional Results: Document "MENU_summ.pdf" contains MENU study summary statistics.

Additional Results: Document "rfh_summ.pdf" contains RfH study summary statistics.

Matlab-code: File "fmatch_subj_local.m" contains example Matlab code for estimating the deformation between baseline and follow-up PA curves (parallel computing for subject id).

R-code: File "code_data_analysis.R" contains example R code for data analysis.

R-code: File "code_sim.R" contains example code for simulations, including comparison with method of extracting PCs from vertical changes in PA only, as well as the two-step approach based on Wrobel (2019).

R-code: File "warp_subj_fn.R" contains a wrapper function to warp baseline to target curve using Wrobel (2019)'s method, and it is called in code_sim.R.

UCSD (MENU), focusing on the effect of interventions on longitudinal changes in PA patterns and how different modes of changes in PA influence weight loss, respectively. The proposed approach reveals unique modes of changes, including overall enhanced PA, boosted morning PA, and shifts of active hours specific to each study cohort. The results bring new insights into the study of longitudinal changes in PA and health and have the potential to facilitate designing of effective health interventions and guidelines.

Keywords

Activity trackers; accelerometer; functional data analysis; Riemann manifold; longitudinal analysis; functional principal component analysis

1. Introduction.

Physical activity (PA) has been linked to health in many epidemiological studies. Physical inactivity and sedentary behavior are known risk factors for cardiovascular disease, cancer, and all-cause mortality (Chastin et al. (2019), LaCroix et al. (2019), LaMonte et al. (2018), Bellettiere et al. (2019a, 2019b, 2020), Glass et al. (2021), Parada et al. (2020), Ramakrishnan et al. (2021), Walker et al. (2021)).

Activity trackers, based on sensor devices such as accelerometers, enable detailed tracking of daily PA at minute-level in free-living environment. In studies using activity trackers, participating subjects wear the device in body parts, such as hip, thigh, or wrist, and perform their usual daily activities nonintrusively (Adamo et al. (2009), Ainsworth et al. (2014), Colley et al. (2011), Dyrstad et al. (2014), Prince et al. (2008)). For systematic reviews and meta-analyses on data from accelerometer-based trackers, see Ekelund et al. (2019), Migueles et al. (2017), Stamatakis et al. (2019).

The emerging usage of activity trackers in studies provides an unprecedented opportunity for in-depth research on physical activity and sedentary behavior. In particular, recording PA of subjects with activity trackers in multiple visit periods, including the baseline and follow-ups, not only reveals information on PA during each period but also enables estimation of longitudinal changes in PA between periods/visits. As a result, the impact of interventions, if incorporated in the study, can be evaluated by examining the association between the intervention and longitudinal changes in PA. Moreover, the effect of longitudinal changes in PA on health outcomes, such as obesity and physiological measurements relevant to diseases, can be studied in correlative analysis of changes in PA and outcomes of interest.

The aforementioned analyses require statistical modeling of PA with minute-level activity tracker data and of the longitudinal changes in PA. Past studies of activity tracker data rely heavily on summarizing minute-level data into single day-level measurements, such as assessed sedentary time (SED), light-intensity physical activity (LIPA), and moderate-to-vigorous physical activity (MVPA) (Füzéki, Engeroff and Banzer (2017), Loprinzi (2016), Matthews et al. (2002), Nader et al. (2008)). Each of the day-level measurements summarizes a facet of the activity tracker data. However, important information can be lost in aggregating minute-level PA records. In particular, the temporal dependence in PA at

different times of a day cannot be assessed after day-level reductions, which prevents the inclusion of PA diurnal patterns in the scope of further studies.

Functional data analysis (FDA) approaches that preserve the original format of the minute-level data instead of aggregating them into summary statistics have been adopted in explaining variability in PA, especially in temporal activity patterns, among subjects within one cohort. In these studies PA is usually characterized as a subject-specific $X_{ij}(t)$ indexed by chronological time t, where i and j index subjects and visits, respectively. Variability in PA is often modeled with functional principal component analysis (fPCA) type of variance decomposition method. For example, Li, Staudenmayer and Carroll (2014) modeled jointly the energy expenditure and interruptions to sedentary behaviors with fPCA. Goldsmith, Zipunnikov and Schrack (2015) proposed a generalized multilevel fPCA model to study minute-level PA recorded by activity trackers of subjects in multiple days. Li et al. (2015) extended the model in Goldsmith, Zipunnikov and Schrack (2015) and proposed a three-level functional data model allowing daily records to be nested in weeks. For details of multilevel fPCA and its applications to longitudinal functional data, see, for example, Di et al. (2009) and Greven et al. (2010, 2011).

Studies focusing on other aspects of PA records and functional data in general have also been conducted. Xiao et al. (2015) proposed a covariate-dependent functional model to quantify the lifetime circadian rhythm of PA. Shou et al. (2015) discussed a structured fPCA to handle multiple levels of variation generated by nested and crossed study designs. Goldsmith et al. (2016) adopted function-on-scalar regression methods to relate 24-hour diurnal PA patterns with covariates. Xu et al. (2019) implemented an fPCA mixed model to study multiple daily PA records and examined the association between activity patterns and health outcomes. Wrobel et al. (2019) introduced an fPCA approach for estimating a template for aligning functions in temporal domain and a time-warping method for registration of subject-specific functions to the template. Choi et al. (2018) employed temporal registration procedures prior to classifications of activity types using machine learning methods. Srivastava et al. (2011) proposed a geometric framework for separating the phase and the amplitude variability in functional data. The registration is based on warping using the Fisher–Rao Metric. Anirudh et al. (2016) approached the problem of functional data registration based on transported square-root velocity fields. Kurtek (2017) developed a Bayesian model for pairwise nonlinear registration of functional data and explored statistical inference tools including k-means for clustering. Reuter et al. (2020) proposed a two-stage clustering method for sedentary behavior (SB) and examined associations between SB patterns and longitudinal physical functioning with a mixed model.

Despite the recent development of FDA approaches for minute-level PA data, it remains unclear how to directly model longitudinal changes in PA diurnal patterns and how to correlate changes in PA with health interventions and outcomes. Motivated by two clinical trials, RfH, and MENU, we propose a new framework by applying Riemann manifold theory to the analysis of longitudinal PA records. The model framework consists of multiple stages. First, preprocessed and smoothed minute-level PA records are characterized as one-dimensional (1D) Riemann manifolds (curves). Next, the longitudinal changes in PA are modeled as deformations between the manifolds/curves. Specifically, the deformations are

modeled by diffeomorphisms, governed by elements in a reproducing kernel Hilbert space, that satisfy minimal-energy constraints (Beg et al. (2005), Charlier, Charon and Trouvé (2017), Vaillant et al. (2004)). Results on diffeomorphisms have been used in matching and registration of medical images such as structural brain fMRI; see, for example, Hernandez, Bossa and Olmos (2009). However, to our best knowledge, this is one of the first studies of longitudinal changes in physical activity utilizing the power of diffeomorphism theory. We further model the variability in longitudinal changes of PA with fPCA. Principal components (PC) that are capable of explaining a majority of the variability are examined as the main modes of changes in PA patterns, and the corresponding projection coefficients (scores) are used to characterize the composition of different modes of variation in subject-specific PA changes. Finally, relations between longitudinal changes in PA diurnal patterns and health outcomes and/or interventions are analyzed via correlative studies of PC projection coefficients.

An important advantage of the proposed approach is that it models not only changes in PA magnitude at each fixed time point of a day in different visit periods but also temporal/phase shifts of activity patterns. The capability of capturing phase changes is critical in the study of PA, as subjects can experience change of circadian rhythms and may exhibit a shift of active hours including changes in sleep/wake up cycles associated with health outcomes and interventions (Khan et al. (2021), Montaruli et al. (2017)). The proposed approach enables identification of such important changes in PA patterns, and results can facilitate the design of effective interventions to promote beneficial PA habits.

The paper is organized as follows. Section 2 introduces the two clinical trials that motivate the proposed approach. Section 3 describes the Riemann manifold model framework. Section 4 delineates the estimation pipeline for parameters in the proposed model. Section 5 conducts simulation studies to evaluate the proposed approach. Sections 6 and 7 conduct comprehensive data analyses on data from the RfH and MENU studies. Technical details and additional simulation and data analysis results are available in the Supplementary Material Zou et al. (2023).

2. Data motivation.

The model and method proposed in this paper are motivated by two clinical trials: Reach for Health (RfH) and Metabolism, Exercise and Nutrition at UCSD (MENU). Both studies were part of the NIH-funded Transdisciplinary Research on Energetics and Cancer (TREC) Study at University of California, San Diego (UCSD) from 2011 to 2017. The RfH study is a six-month clinical trial that involved 333 overweight, postmenopausal early-stage breast cancer survivors. Each of the subjects was randomly assigned to either metformin treatment or placebo, combined with either a lifestyle-based intervention or placebo. The primary aim was to test the effect of treatments on health outcomes, especially weight loss, as measured by the change in BMI (Patterson et al. (2018)). The MENU study is a 12-month behavioral intervention study on 245 overweight and otherwise healthy women. Each subject was randomly assigned to one of the three diet treatment groups: a lower fat (20% of energy) and higher carbohydrate (65% of energy) diet, a lower carbohydrate (45% energy) and higher monounsaturated fat (35% energy) diet, or a walnut-rich (35% fat) and lower carbohydrate

(45%) diet (Le et al. (2016), Patterson et al. (2016), Rock et al. (2016)). The goal was to study the role of dietary macronutrient composition on weight loss.

In both studies multiday PA at baseline and follow-up periods were measured by the GT3X ActiGraph (ActiGraph, LLC; Pensacola, FL; www.actigraphcorp.com), a research-grade activity tracker with built-in triaxial accelerometer. Activity data were collected at high-resolution of 30 Hz, then processed into per minute PA vector magnitudes (VM, the L^2 norm of the triaxial PA records) (Bassett (2012), Xu et al. (2019)). Our goal is to characterize longitudinal changes in PA diurnal patterns and examine the relations between changes in PA patterns and health outcome (for the MENU study) and interventions (for the RfH study).

3. The model framework.

3.1. A Riemann manifold model for physical activity.

Suppose there are N subjects in the study. For the ith subject, daily PA is recorded by the activity tracker in periods $k = 0, \ldots, K$, where $k = 0$ denotes the baseline visit and $k = 1, \ldots, K$ denote the follow-ups. The number of visits is assumed to be the same for all subjects. Data of each period consist of daily minute-level PA VM records of several consecutive days. The multiday PA records in each period are averaged at each of the chronological time t of a day and smoothed to reduce noise (details in Section 4). For the ith subject, the averaged and smoothed PA measured at the kth period in the longitudinal study is modeled as a one-dimensional Riemann manifold/curve $X_i^{(k)}$ with image on $[0, T] \times [0, M] \subset \mathbb{R}^2$, where $t \in [0, T]$ is the index for chronological time of a day and M denotes the maximal admissible PA magnitude for all subjects.

For subject i we characterize the longitudinal change in the PA from visit $(k-1)$ to k as

$$X_i^{(k)} = \phi(\boldsymbol{v}_i^{(k)}, \cdot) \circ X_i^{(k-1)} = \phi(\boldsymbol{v}_i^{(k)}, X_i^{(k-1)}), \tag{1}$$

where $\phi(\boldsymbol{v}_i^{(k)}, \cdot)$ is a subject-specific diffeomorphism that maps the PA curve $X_i^{(k-1)}$ at the $(k-1)$th period to the PA curve $X_i^{(k)}$ at the kth period. The symbol $\circ$ denotes the composite of two functions. The diffeomorphism is assumed to be governed by $\boldsymbol{v}_i^{(k)}$, which is a subject-specific element in $L^2([0,1], V)$. The space V contains vector fields defined on $[0, T] \times [0, M]$, and each element $\boldsymbol{v} \in L^2([0,1], V)$ is a set/series of time-varying vector fields, indexed by $\tau \in [0,1]$, with finite L^2 norm: $\int_0^1 \|v_\tau\|^2 d\tau < \infty$, where v_τ denotes the value of $\boldsymbol{v}$ at a fixed index τ. Note τ should not be mixed up with the chronological time t in a day.

The diffeomorphism $\phi(\boldsymbol{v}_i^{(k)}, \cdot)$ in (1) is obtained by *flowing* $\boldsymbol{v}_i^{(k)}$. That is, for any $\boldsymbol{v} \in L^2([0,1], V)$, let $\psi_{\boldsymbol{v}}$ be the solution to the ordinary differential equations $\frac{\partial}{\partial \tau}\psi_{\boldsymbol{v}}(\tau, \cdot) = v_\tau \circ \psi_{\boldsymbol{v}}(\tau, \cdot)$ with the initial condition that $\psi_{\boldsymbol{v}}(0, \cdot)$ is the identity mapping. The deformation $\phi(\boldsymbol{v}, \cdot) = \psi_{\boldsymbol{v}}(1, \cdot)$ is defined to be the final state of $\psi_{\boldsymbol{v}}$. Intuitively, $\boldsymbol{v}$ can be considered as a series of forces that "drag" the PA curve from one period to another at each

$\tau \in [0, 1]$. At each step τ, the force drags the current PA curve by the amount of v_τ until the final step $\tau = 1$. Equation (1) states that, with the subject-specific vector field $\boldsymbol{v}_i^{(k)}$, the total deformation applied to the PA curve at period $(k-1)$ is $\phi(\boldsymbol{v}_i^{(k)}, \cdot) = \psi_{\boldsymbol{v}_i^{(k)}}(1, \cdot)$ to obtain the PA curve at period k. For technical details on the diffeomorphisms, see Charlier, Charon and Trouvé (2017) and the references therein.

The $\boldsymbol{v}_i^{(k)}$ at each $\tau \in [0, 1]$ is further assumed to follow a reproducing kernel Hilbert space (RKHS) representation:

$$v_\tau(\cdot) = \sum_{j=1}^{n_g} K_V(\psi_{\boldsymbol{v}}(\tau, c_j), \cdot) m_{j,\tau}, \tag{2}$$

where $\{c_j : j = 1, \ldots, n_g\} \subset \mathbb{R}^2$ is a set of preselected control points of total number n_g. $K_V(x, y) = \exp(-\|x-y\|^2 / (2\sigma_V^2))\mathbf{I}_2$ is the Gaussian isotropic kernel with a fixed rigidity parameter σ_V^2, and $m_{j,\tau} \in \mathbb{R}^2$ is the momentum of the deformation at the jth control point at τ. Intuitively, (2) states the force that drags the PA curve is determined by forces (momenta) on a finite number of points on the curve, for example, at a fixed number of minutes in a day. The dragging forces applied to noncontrol points can be represented as weighted averages of those on the control points.

Given $X_i^{(k-1)}$ and $X_i^{(k)}$, the deformation $\phi(\boldsymbol{v}_i^{(k)}, \cdot)$ that satisfies (1) is not unique. Further constraints are needed to uniquely define the vector field $\boldsymbol{v}_i^{(k)}$ and its associated deformation $\phi(\boldsymbol{v}_i^{(k)}, \cdot)$. Here we adopt the constraint on the deformation energy $\int_0^1 \|v_\tau\|^2 d\tau$. We model the change in the ith subject's PA between the $(k-1)$ and the kth periods as the deformation $\phi(\boldsymbol{v}_i^{(k)}, \cdot)$ that satisfies (1) with $\boldsymbol{v}_i^{(k)}$ minimizing the energy. Subject to the constraint on the deformation energy, the $\boldsymbol{v}_i^{(k)}$ that satisfies (1) is unique and is determined solely by the values of $\boldsymbol{v}_i^{(k)}$ at $\tau = 0$ (Charlier, Charon and Trouvé (2017)).

With the above model, the variability in the longitudinal changes in PA among subjects in the cohort can be characterized by the subject-specific random vector fields $\{\boldsymbol{v}_i^{(k)} : i = 1, \ldots, N, k = 1, \ldots, K\}$. Furthermore, with the RKHS representation (2), the *initial momenta* $(m_{1,0}^i, \ldots, m_{n_g,0}^i)$ (initial meaning at $\tau = 0$) fully determine the deformation, and we denote the initial momenta by m^i when there is no ambiguity.

3.2. Functional principal component analysis (fPCA) for the deformations.

The variability in longitudinal changes in PA among subjects in a cohort can be represented further by a function principal component model. For simplicity, we first assume $K = 1$, which means there is only one follow-up after the initial (baseline) visit/period. The case of multiple follow-ups can be readily extended: an extra dimension can be introduced for the visits so that the domain of the vector fields at each τ becomes $\{1, \ldots, K\} \times E$, and for each $k = 1, \ldots, K$, the sub-vector-field $v : \{k\} \times E \to E$ denotes the vector field governing the

deformation from the PA curve at the $(k-1)$ visit to the kth visit. Then the same analysis can be applied to the extended-dimension vector fields.

Consider the principal component representation of the initial momenta $m^i = \bar{m} + \sum_{l=1}^{\infty} a_{il}\mu_l$, where $\bar{m}$ is the mean of the momenta and $\{\mu_l : l = 1, \ldots, \infty\} \subset \mathbb{R}^{n_g \times 2}$ are the principal components (PC). The $a_{il} = \langle m^i - \bar{m}, \mu_l \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the Frobenius inner product on $\mathbb{R}^{n_g \times 2}$, is the projection coefficient (score) of the mean-subtracted m^i on the lth PC. Assuming the variability in the deformations of subjects in the cohort can be well explained by a finite number N_{pc} of principal components, we can write approximately $m^i = \bar{m} + \sum_{l=1}^{N_{\text{pc}}} a_{il}\mu_l$. In practice, the value of N_{pc} can be determined by researchers based on criteria including the proportion of variance explained by the top PCs.

3.3. Correlative study with health outcomes.

Two types of questions are often of interest to researchers: 1. how do PA diurnal patterns change as responding to interventions designed for encouraging more active lifestyle, and 2. how do longitudinal changes in PA affect health outcomes? To address these two questions, we model the relation between changes in PA and intervention/health outcomes with regressions in which the longitudinal changes in PA are modeled via the PC projection coefficients/scores $\{a_{il}\}$. The model is $a_{il} = f_l(W_i, Z_i) + \epsilon_{il}$ for effect of intervention W_i and covariates Z_i on changes in PA. Here the regression is carried out separately for each PC indexed by l. The model is $Y_i = f(a_{i1}, \ldots, a_{iN_{\text{pc}}}, Z_i) + \epsilon_i$ for effects of changes in PA on health outcome(s) Y. In both models the ϵ are independent and identically distributed (i.i.d.) errors.

4. Estimation methods.

In this section we delineate the estimation procedure for model parameters in Section 3. We focus on two groups of parameters: the parameters in modeling the longitudinal changes in PA via diffeomorphisms and the parameters in representing the group-level variability of changes in PA with the fPCA model. Parameters in models in Section 3.3 can be estimated with standard methods for regressions.

4.1. Estimation of diffeomorphisms and initial momenta.

In this section we focus on the estimation of the deformation vector fields $\{\boldsymbol{v}_i^{(k)}\}$ and its associated initial momenta $\{m^i\}$ in the diffeomorphism model described in Section 3.1. The flowchart in Figure 1 summarizes the steps involved; in what follows, we introduce the details in each step.

In practice, each subject's PA measured by the accelerometers of activity trackers are available to the researcher in the discrete format. In each visit/period, there are usually several consecutive days of PA records. For each day, preprocessed PA records are available in every minute of a day. For each subject the multiday PA records in the same period are further averaged to reduce noise.

The averaged PA records in each visit are then smoothed using splines with function *smooth.spline()* in software R 4.0.3 (R Core Team (2021)) to further reduce noise caused by minor fluctuations of the activity tracker that are irrelevant to meaningful movements. We choose the degree of freedom (trace of smoothing matrix) $df = 25$ in spline smoothing based on exploratory analysis of datasets from both RfH and MENU studies so that the smoothed PA records preserve visible diurnal PA patterns while reducing noise induced by the trackers. Note here we use the same degree of freedom in smoothing the RfH and MENU PA records, as both studies use the GT3X Actigraph activity tracker to measure subjects' PA, leading to comparable noise levels in PA records. It is also important to use consistent smoothing and scaling parameters for all subjects' PA at all periods to preserve informative variability in the PA. Figure 2 displays the smoothed PA curves at baseline visiting period and month 6 from an example subject in the RfH study. A sensitivity analysis on the smoothing parameter is available in the Supplementary Material in which multiple values of the smoothing parameters are examined and options are discussed for practical use of the smoothing procedure. In summary, the proposed approach is generally robust when the smoothing parameter is in a reasonable range. When smoothing is not adequate, the remaining noise can influence the results, especially estimated PCs that explain less variability of the data. For a different study using other activity tracker devices, it is recommended to conduct exploratory analysis and/or lab calibrations of the data to determine the appropriate smoothing parameter.

For each subject i and period k, we estimate the subject-specific deformation between smoothed PA curves $X_i^{(k-1)}$ and $X_i^{(k)}$. The estimation of the deformation that transforms a subject's PA curves from one period (the *source*) to another (the *target*) is formulated as an optimization problem: the goal is to find the deformation that minimizes a dissimilarity metric g between the deformed source curve and the target curve, while constraining on the total energy $\int_0^1 \|v_\tau\|^2 d\tau$ of the deformation. In computing the optimal deformation, the optimization problem with constraint on energy is transformed into an unconstrained problem with the following penalized objective function:

$$\frac{\gamma_W}{2} g\left(\widetilde{X}_i^{(k)}, X_i^{(k)}\right) + \frac{\gamma_V}{2} \int_0^1 \|v_\tau\|^2 d\tau, \tag{3}$$

where g is the norm of the difference between the observed target curve $X_i^{(k)}$ and the estimated target curve $\widetilde{X}_i^{(k)} = \phi(\boldsymbol{v}_i^{(k)}, \cdot) \circ X_i^{(k-1)}$ resulting from deforming $X_i^{(k-1)}$ according to vector fields $\boldsymbol{v}_i^{(k)}$. The norm of the difference is defined based on Gaussian kernels between two discretized curves (for details, see equation (62) of Charlier, Charon and Trouvé (2017)). Here γ_V and γ_W are preselected penalty parameters for the energy and dissimilarity terms. The goal is to find the optimal subject-specific vector fields $\boldsymbol{v}_i^{(k)}$ for (3).

As discussed previously, the vector fields $\boldsymbol{v}_i^{(k)}$ are assumed to follow an RKHS representation (2) and are fully determined by the initial momenta m^i at a set of preselected control points. Therefore, optimizing the objective (3) over admissible values of $\boldsymbol{v}_i^{(k)}$ is equivalent to finding

the optimal initial momenta for the objective function. A natural choice of the control points is the set of available time points (minutes) and associated PA magnitudes in the data. The optimization is conducted using the Hamiltonian method with the function "fsmatch_tan()" in the package *fshapesTK* (Charlier, Charon and Trouvé (2017)). An example Matlab script to estimate subject-specific deformation is available in the Supplementary Material.

As an example, Figure 3 shows the estimated deformation and initial momenta of one subject in the RfH study. Arrows illustrate the initial momenta at control points (arrows are drawn every 10 minutes for a better visualization). The initial momenta capture several key characteristics of changes in PA patterns. PA levels increase from baseline to month 6, most significantly, in time windows 6 a.m. to 12 p.m. and 1 to 3 p.m., reflected by the upward pointing arrows. The arrows of momenta also signal two visible shifts of active hours: the small peak, centered around 9 a.m. at baseline visit, shifts slightly to the right after six months, while the peak, centered around 12 p.m., shifts significantly to the left.

Note that the second term in the objective function (3) is a penalty on the deformation energy $\int_0^1 \|v_\tau\|^2 d\tau$, which imposes a penalty on the initial momenta m^i with equal weights on its x and y coordinates. An alternative approach is to impose different weights on x and y coordinates. How to choose the weights, however, is a substantive question, determined by the emphasis on the temporal (x) and vertical (y) changes of the PA. Alternatively, one can apply scaling to x and y coordinates of the PA curves to control the ranges of both coordinates. A smaller range results in less energy consumption and thus is equivalent to less penalty weight. In our analysis for the RfH and MENU data, we adopt the latter approach and center and scale the time indices and PA magnitudes for them to be in comparable scales so that the temporal shift and vertical changes of PA patterns are emphasized equally.

4.2. fPCA of momenta.

The estimated initial momenta for the ith subject $\hat{m}^i := (\hat{m}^i_{1,0}, \ldots, \hat{m}^i_{n_g,0})$ is a vector field defined on the n_g control points and stored in an $n_g \times 2$ matrix. At the lth control point, $\hat{m}^i_l$ is a vector in $\mathbb{R}^2$ indicating directions and magnitudes of the momenta in both the temporal (x-axis) and vertical directions (y-axis). The y direction indicates changes in PA levels, while the x direction indicates temporal shifts of PA patterns.

Suppose a finite number of functional principal components (PC) are adequate in explaining the majority of the variability in the deformation from the baseline to the follow-up PA curves for all subjects. Calculating the PCs of dimension $\mathbb{R}^{n_g \times 2}$ under the Frobenius inner product is equivalent to calculating the PCs of vectors obtained from concatenating the $\hat{m}^i$ matrices. Note the mean of the estimated initial momenta needs to be subtracted prior to estimating the PCs. To interpret and visualize the PCs, we reshape the estimated initial momenta for each PC to the original dimensionality $n_g \times 2$, and "flow" (as described in Section 3.1) a preselected template curve with the deformation governed by the estimated initial momenta using the Matlab function "shoot_and_flow_tan()" in package *fshapesTK* (Charlier, Charon and Trouvé (2017)).

5. Simulation studies.

In this section we evaluate the proposed approach with a synthetic dataset and compare the proposed approach to two existing methods. First, we generate a smooth baseline curve using the averaged baseline PA of the MENU study data, which consists of PA vector magnitudes (VM) at each minute from 7 a.m. to 9 p.m. Here we use the MENU study data to generate the synthetic data in order to reflect the realistic PA patterns in the simulations, and the time window of 7 a.m. to 9 p.m. is selected for demonstrative purposes only. Then we create three principal components (PC) by first manually creating three sets of initial momenta and then running functional PCA (with function *FPCA()* of R package *fdapace* by Gajardo et al. (2021)) to extract orthogonal eigen-functions of them. The resulting mutual-orthogonal initial momenta represent three modes of changes in PA diurnal patterns from the baseline to the follow-up visit. Figure 4 shows the initial momenta of the PCs as well as the follow-up curves resulting from applying the PCs on the mean baseline PA curve. The first PC is a general increase of PA levels throughout the day, the second is a local boost of PA, and the third is mainly a time shift of active hours to later times of the day.

We generate each subject's actual initial momenta of deformations as linear combinations of the initial momenta of the three PCs, where the coefficients in the linear combinations are simulated from independent Gaussian distributions with mean 0 and standard deviation l, scaled by (0.8, 1.2, 1.5)/2000 for the coefficients of PC 1, 2, and 3, respectively, to make a distinction between the PCs. Each subject's follow-up PA curve is generated by deforming the baseline PA curve with the subject-specific initial momenta following the procedure described in Section 3. Figure 5 visualizes an example subject's baseline and follow-up PA curves in which the weights are 40%, 15%, and 45% for PC 1 (with negative sign), 2, and 3. The follow-up curve deformed from the baseline exhibits an apparent time shift, as in PC 3, as well as a PA magnitude change throughout the day as in PC 1. The local PA magnitude change in PC 2 is less manifested due to the smaller weight.

With the observed baseline and follow-up PA curves, we estimate the subject-specific initial momenta of the deformation with the proposed method. As a comparison, we also calculate the difference between the follow-up and baseline PA magnitudes at each minute and extract PCs only from the vertical differences in PA magnitudes. Interpolations using the R function *approx()* are applied to time points where follow-up PA magnitudes are missing due to the deformation in the temporal domain. In addition, we include in the comparison a two-step approach based on the time-warping method in Wrobel et al. (2019). In the first step, each subject's baseline curve is registered to the follow-up curve using the package *registr2.0* (Wrobel et al. (2022)). Necessary formatting and standardization are applied to the curves prior to applying the package. The output of the registration for each subject is the set of warped time indices as a function of the original time indices in the baseline curve. Then functional PCs are extracted from the warped time indices using the *fdapace* package. In the second step, differences in the PA magnitude between the follow-up curves and the warped baseline curves are calculated and PCs are extracted from the differences.

Figures 6 and 7 show the estimated PCs using the proposed method and the PCs extracted from vertical differences. The proposed method to a large extend recovers the modes

of longitudinal changes in PA patterns by capturing most of the vertical changes in PA magnitudes and temporal shifts of PA patterns. On the other hand, analysis of only the vertical changes between the baseline and follow-up PA magnitudes is unsuccessful in revealing the major changes in PA diurnal patterns: the overall increase of PA in the actual PC1 is only partially recovered with missing increase patterns around 10 a.m. to 12 p.m.; the period of locally boosted PA in PC2 is incorrectly extended to 9 a.m. to 10 a.m. The temporal shift in PC3 is almost completely overlooked. The estimated PC 3 tries to recover the shift by having a local decrease before 10 a.m. and a local increase around 10 a.m. to 1p.m. but was unsuccessful in capturing the shift of PA patterns, as made clear by the dashed curves. Correlations between the estimated and actual PC momenta and scores are available in the Supplementary Material.

Figure 8 shows the PCs estimated from the two-step approach based on the time-warping method in Wrobel et al. (2019). The bottom panel shows the estimated follow-up curve when applying the PC 1 extracted from the time-warping functions, which explained over 97% of the total variance in the subjects. Note, there are no momenta involved in this method, and thus no arrows are plotted. Comparing to the actual PC 3 in Figure 4, the estimated PC captures most of the shift between 9 a.m. and 1 p.m. but tends to overestimate the range of the shift: estimated shifts are also visible from 7 a.m. to 9 a.m. and 1 p.m. to 3 p.m. The top and middle panels show the top two PCs extracted from the vertical difference between the follow-up PA curves and the time-warped baseline curves. Comparing to the actual PC 1 in Figure 4 of an overall increased/decreases change in PA magnitude between 9 a.m. and 5 p.m., the estimated PC 1 is capable of recovering the change in the time window of 10 a.m. to 1p.m. but fails to capture the changes outside this range. The estimated PC 2 resembles the actual PC 2 in the local change of PA magnitude between 10 a.m. to 11 a.m., but the estimated change is less apparent comparing to the actual PC, especially when contrasting with the change in the opposite direction in the time window of 11:30 a.m. to 3 p.m. There are also some mix-ups in the PC 1 and PC 2 of the PA magnitude change. The difference in results can be partially due to the sensitivity of the estimation of warping to the change in PA magnitude, especially to changes in small/local time windows as those in PC 2. While the an apparent overall increase/decrease in PA magnitude between the baseline and the follow-up curves can be adjusted via centering and standardization prior to applying the warping estimation, changes in local time windows are more difficult to adjust and can lead to a mix-up in the estimation of time warping and vertical change in PA magnitude. On the other hand, the proposed approach is capable of accounting for both changes in the temporal domain and in the vertical PA magnitude and achieve more accurate estimation of both changes.

6. Data analysis I: The RfH study.

In this section we analyze the RfH study data, which consist of PA records at baseline and month-6 for 333 overweight, postmenopausal early-stage breast cancer survivors. Each subject was randomly assigned to one of the 2 × 2 lifestyle intervention and medication treatment arms. Our focus here is to study the effect of the treatments on the longitudinal changes in PA.

6.1. Preprocessing of data.

Subjects with both baseline and month-6 measurements available from the activity tracker are retained for the analysis. As a result, 303 out of 333 subjects remain in the dataset after cleaning for missing records. For each subject the baseline and month-6 activity tracker measured PA records are sorted by dates and time of wearing. We only use PA records of each day from 6 a.m. to midnight, as most of the records out of this range can be omitted due to nonwear time when the subject was sleeping. Within the 6 a.m. to midnight window, whenever the subject was not wearing the device, PA was recorded as zero and treated as inactive time due to sleeping, since subjects were asked to wear the activity tracker during all waking hours, except for time contacting water (Patterson et al. (2018), Xu et al. (2019)). Note here we use the chronological time to index the PA records instead of align subjects' PA records, according to the subject-specific time of first nonzero record, as of interest here are the diurnal activity patterns as a function of chronological time and the longitudinal change in such activity patterns.

The retained PA records for each subject are then averaged over multiple days of measurements within each period, and the resulting records have format $(t, \mathrm{PA}(t))$ for each period, where $t = 1, \ldots, 2160$ are minutes from 6 a.m. to midnight. Smoothing with splines is then applied to the averaged baseline and month-6 activity PA to get the smoothed PA curves. The smoothing reduces the variability induced by random measurement errors and fluctuations caused by the device and to increase the signal-to-noise ratio in the activity record to avoid overfitting in the subsequent analyses. The same smoothing parameters are used for all baseline and month-6 PA records for all subjects. The smoothing parameter $df = 25$ is selected based on exploratory analysis of both RfH and MENU datasets. Finally, we center and scale the time and PA magnitudes so that they are both in the range $[-1, 1]$ approximately. As discussed in Section 4.1, this step balances the penalties on the deformation energies of the initial momenta on both x and y coordinates.

In addition to PA, demographic and medical measurements were collected for each subject in both baseline and month-6 visits. In particular, treatment assignment indicators of medication (metformin and placebo) and life-style intervention (with and without weight-loss program) are available both separately and as a 2×2 factorial design. Other important variables include demographics, body mass index (BMI), activity minutes at baseline, history of diseases and medication, pathological stage, years from diagnosis, and glucose.

6.2. Estimating longitudinal changes in PA.

The first step of the estimation is to characterize the longitudinal changes of PA in terms of subject-specific deformations from the baseline activity curves to the month-6 curves. As discussed in Section 3, each subject's deformation from the baseline curve to the month-6 curve is modeled as a diffeomorphism governed by a vector field, and the vector field is assumed to follow the RKHS representation (2). To estimate the deformations, it suffices to estimate the initial momenta at the control points.

Here we use the function *fsmatch_tan()* in the Matlab package *fshapesTK* (Charlier, Charon and Trouvé (2017)) to estimate the subject-specific initial momenta in deforming the

baseline to the month-6 PA curve. Recall the deformation at each middle step $\tau \in (0, 1)$ follows the differential equation $\partial\psi_v(\tau, \cdot) / \partial\tau = v_\tau \circ \psi_v(\tau, \cdot)$. In estimating the momenta, we discretize τ into finite values $\{\tau_k : k = 1, \ldots, N_{\text{steps}}\}$ and estimate each v_{τ_k} in each step. Upon examining the discretization, we find $N_{\text{steps}} = 11$ steps lead to consistently satisfactory approximation of the deformation. Figure 3 shows estimated deformation of an example subject including the deformed curve at $\tau = 4 / 11$. For data of much higher dimensionality, exploratory analysis can be conducted to find the appropriate number of steps in the discretization. It takes about two minutes to estimate the deformation from the baseline to the follow-up PA curve for each subject. High computing cluster is used to run the estimating procedure for all subjects in parallel. Details on function parameters in the estimation are available in the Supplementary Material.

6.3. fPCA estimation.

The estimated initial momenta of the deformations are concatenated into vectors, and the principal components (PCs) of the vectors are extracted from the mean-subtracted momenta using the function *FPCA()* of R package *fdapace* (Gajardo et al. (2021)). The top 30 explain about 70% of the variance in the data. In what follows, we will focus on the analysis involving the top 10 PCs that explain about 40% of the variance. Other PCs that explain less variability in the data and results are deferred to the Supplementary Material.

6.4. Intervention and covariate associations with PA change.

We are interested here in the effects of the covariates, the interventions on activity, in particular on different modes of longitudinal changes in PA. To this end we conduct an analysis on each of the top 10 PCs. In each separate analysis, the projections (scores) of subject-specific deformations on each PC is the outcome of interest and life-style intervention as well as relevant demographic and medical variables are included as covariates. Table “RfH Summary Statistics” in the Supplementary Material lists summary statistics of the interventions and other variables. Variables of with more than 5% values missing are discarded for further analysis. Medical conditions with less than 5% missing are imputed with “no” for the missing values.

Lasso regularization with 10-fold cross-validation for the choice of penalty parameter is used to select variables from the covariates. Several variables are discovered as significantly associated with modes of change in PA, as characterized by certain PCs. In what follows, we focus on results of PC 1 and 3, which are found to be associated with lifestyle intervention. Significant effects of covariates on the corresponding PC projection coefficients are listed, and initial momenta of PCs in question are visualized along with the estimated month-6 curve resulting from deforming the mean baseline curve of all subjects according to the diffeomorphism governed by the initial momenta of the PC. For the purpose of demonstration, we also visualize the month-6 curve resulting from deforming the baseline with the estimated initial momenta of the PC multiplied by −1 to illustrate the opposite direction of the PA change if a subject has negative projection coefficient on the PC. Note each PC can be scaled by a positive or negative constant and still be a well-defined PC as long as the PC scores change are scaled accordingly. For simplicity in interpreting the

results, we scale the PCs and corresponding projections/scores so that subjects' projections of each PC have standard deviation equal to one.

6.4.1. ***PC* 1.—**Figure 9 visualizes the first PC. As PC 1 explains most variability in the data, it is not surprising that the mode of change in PA characterized by PC 1 is an overall decrease of activity level from the baseline to month 6 (or overall increase if consider $-1 \times \text{PC } 1$). In particular, activity levels in the morning from around 9 a.m. to 5 p.m. change most significantly, as illustrated by the downward pointing arrows.

In modeling effects on PC 1 while adjusting for other covariates, lifestyle intervention is selected to be included in the model with Lasso regularization. The Lasso penalty parameter is chosen based on a 10-fold cross-validation. Table 1 shows the result of a rerun of a separate unregularized regression analysis using only variables selected by Lasso regularization. Lifestyle intervention is significant with a negative effect −0.265 and hence a positive effect on $-1 \times \text{PC } 1$, with p-value 0.015, indicating the intervention potentially has a positive effect on promoting daily activity.

Table 1 also lists estimated effects of Lasso-selected covariates on overall decrease in PA, as characterized by a positive projection score on PC 1. Cut-point based average minutes categories at baseline: LPA100 (Light Physical Activity with CPM, count per minute of the VM, in the range of 100 – 1951), and MVPA1952 (Moderate Vigorous Physical Activity with CPM > 1952) are significantly correlated with PC 1. For details on above cut-point based categories, see, for example, Amagasa et al. (2018). Signs of the effects associated with the above categories (positive for both LPA100 and MVPA1952) are reasonable, as those who are less active and more sedentary are expected to have more space of improvement and thus exhibit an increase in PA during the study. Additional diagnostics plots are available in the Supplementary Material.

Subjects with diabetes tend to exhibit increased activity. Given the relatively small number of subjects with diabetes (seven out of 303), however, the effect of diabetes needs to be validated in a more comprehensive analysis of a more balanced dataset. Taking aspirin (such as Anacin, Bufferin, Bayer, Excedrin) regularly more than three times per week at baseline visit is also associated with increased activity. Employment status of disabled and/or retired because of health is associated with decreased PA comparing to other employment status groups. Lastly, the randomized assignment of medication (metformin or placebo) is not selected by Lasso.

6.4.2. ***PC* 3.—**Figure 10 visualizes the estimated PC 3. The mode of variation in PA associated with PC 3 is marked by a mild increase of PA level at midday (10 a.m. to 2 p.m.) and a decrease in the evening (6 to 9 p.m.). In addition, there is a possible shift of PA pattern to early hours at around 6 to 8 p.m. and a less manifested shift to later hours in the morning between 6 to 9 a.m.

Table 2 summarizes estimated effects of Lasso-selected covariates in the rerun of regression for PC 3. Life style intervention has a significantly negative effect, indicating subjects in the intervention group are more likely to exhibit a change in PA in the direction of $-1 \times \text{PC } 3$

(as illustrated by the dashed line in Figure 10), with lower PA levels between 9 a.m. to 2 p.m. but higher PA after 6 p.m. and a shift of PA pattern to later hours. Again, randomized assignment of medication (metformin or placebo) is not selected by Lasso as a relevant predictor. Moreover, being a former smoker ($n = 135$) is associated with aforementioned mode of change in PA patterns, comparing to those who never smoke ($n = 164$) or current smokers ($n = 4$). The same phenomenon is observed for employment status of disabled and/or retired because of health. Wear time is positively associated with PC 3, indicating those who wore the device longer tend to exhibit increased PA levels in the middle of the day but decreased PA in the evening.

7. Data analysis II: The MENU study.

The MENU study data consist of 245 overweight nondiabetic women to study the effect of randomly assigned diet interventions on changes in health outcomes in 12 months. Of particular interest is the effect of the diet interventions on weight loss. In line with this aim, for the analysis of the MENU data, we focus on the role of longitudinal changes in PA on weight loss characterized by changes in BMI.

A total of 177 subjects with both baseline and month-12 PA records available were retained for the analysis after processing. Upon observing the raw PA data, only PA records within the range of 6 a.m. to midnight are kept for the analysis as data out of this range were mostly inactive records. The same preprocessing procedures are applied to the MENU study data as for the RfH data. The same estimation pipeline for the subject-specific deformations between the baseline and month-12 PA curves used for the RfH data is applied to the MENU data. fPCA is applied to the estimated initial momenta for each subject's deformation from the baseline PA curve to the month-12 curve with R package *fdapace* (Gajardo et al. (2021)). Top 30 PCs explain over 86% of the variance.

The outcome of interest here is the change in BMI, that is, month-12 minus baseline BMI, as an indicator for weight loss during the 12-month study period. We examine models in which change in BMI is the outcome and subjects' projection coefficients on the PCs are the main covariates of interest. Other important covariates, including diet interventions, demographics, smoking, medical history, and baseline activity metrics, defined by cut-points are adjusted in the model. Table "MENU Summary Statistics" in the Supplementary Material lists summary statistics of the interventions and other variables.

The analysis consists of two steps. First, we fit a regression with Lasso regularization to select variables including the PCs and other covariates that are associated with the outcome. Second, we fit a separate regression, including only Lasso-selected variables as well as race, marital status, and diet intervention, which are also of interest. Table 3 lists the result of regression in the second step of analysis. PCs 1, 10, 20, 25 are discovered to be significantly associated with change in BMI. Model diagnostics plots are also available in the Supplementary Material. Figures 11 and 12 illustrate the estimated deformations associated with PCs 1 and 10. Plots of PCs 20 and 25 are deferred to the Supplementary Material, as they explain a smaller portion of the variance.

Interestingly, PC 1 of MENU is not an overall increase/decrease of PA throughout the day, as discovered in the RfH study. In fact, upon examining the top PCs, none of them exhibits an overall increase of PA (plots of top five PCs are available in the Supplementary Material). Instead, they either display a local increase, as seen in PC 1, or a redistribution of PA throughout the day. The dashed curve in Figure 11 shows that the mode of variation characterized by PC 1 is a local increase of activity in the morning between 6 to 10 a.m., as shown by the upward pointing arrows. The deformed curve from the baseline, following the PC's initial momenta, is illustrated by the solid curve (PC multiplied by −1, as illustrated by the dashed curve). The effect of PC 1 on the change in BMI is negative (−0.507 with p-value 0.01), indicating an increase of morning PA from baseline to month-12 is associated with a decreased BMI.

Figure 12 shows the mode of change in PA associated with PC 10, which exhibits a redistribution of PA. Activity levels are increased significantly around 12 to 2 p.m., 3 p.m. to 5:30 p.m., and 9 to 11 p.m., while decreased slightly around 6 to 7 p.m. and 8 to 9 p.m. There is also a slight shift of PA toward later hours in the early morning shown by the arrows pointing to the right at time window 6 to 8 a.m. The effect of PC 10 on the change in BMI is also negative (−0.428 with p-value 0.027). This result is not surprising, as PC 10 exhibits mostly increased PA in different time windows of a day and thus is expected to be associated with decreased BMI.

Diet interventions are significantly associated with change in BMI. Comparing to the baseline of lower carbohydrate and higher monounsaturated fat diet, the lower fat and higher carbohydrate diet has a negative effect on change in BMI (p-value 0.02). The Walnut-rich diet also has a mildly negative (p-value 0.123) effect. In addition, high cholesterol is significantly associated with positive change in BMI. Larger pain subscale indicates less pain and is associated with decreased BMI. Higher follicle stimulating hormone (FSH) level is associated with decreased BMI. Comparing to the baseline of married, marital status of single or widowed are both associated with decreased BMI.

7.1. Alternative approach with functional regression.

In addition to the functional principal component analysis, we also examine the effect of change in PA diurnal patterns on changes in BMI with the functional regression approach.

The regression coefficients are estimated with methods proposed by Goldsmith et al. (2011). The estimated initial momenta of all subjects in the x (temporal) and y (vertical) coordinates are treated as functional predictors of interest, and the change in BMI is the one-dimensional outcome. For each subject the initial momenta of deformation between the baseline and month-12 PA curves is a vector of 1080 elements in both the x and y coordinates, indicating the directions and magnitudes of changes in PA at each of the minutes between 6 a.m. and midnight. Formally, the model is $\text{bmi}_{\text{month 12}} - \text{bmi}_{\text{baseline}} = \text{intercept} + \alpha Z + \int_t \beta_x(t) m_x(t) dt + \int_t \beta_y(t) m_y(t) dt + \epsilon$, where $(m_x(t), m_y(t))$ denote the initial momenta of deformation in PA at time t for a generic subject, $(\beta_x(\cdot), \beta_y(\cdot))$ denotes the corresponding functional coefficients/effects, Z and α denote time-invariant covariates and corresponding effects, and ϵ denotes the error.

Model parameters are estimated with R package *funreg* by Dziak and Shiyko (2021). We examine the effects while adjusting for important covariates and potential confounders selected by the Lasso model in the previous section. Figure 13 shows the estimates of $\beta_x(\cdot)$ (left panel) and $\beta_y(\cdot)$ (right panel) with 95% confidence bounds. The effect of temporal shift is not significant. The effect of vertical change in PA magnitudes are significantly negative between 8 a.m. to 9 p.m., indicating increased PA in this time window is associated with decreased BMI. For detailed estimated effects of all covariates in the model, see the Supplementary Material. Comparing to results of fPCA, the functional regression result indicates an overall benefit of enhanced PA levels throughout the day (8 a.m. to 9 p.m.) without providing information on detailed PA patterns.

8. Concluding remarks.

In this paper we propose a new framework for longitudinal changes in physical activity (PA) and relations between longitudinal changes in PA and health outcomes as well as interventions. The model and method proposed here are based on a Riemann manifold representation of spline-smoothed PA records as functions of chronological time of a day. The longitudinal changes in PA during a multimonth study period are modeled as deformations between PA curves (1D Riemann manifolds) measured in different visiting periods of the study. The deformations are modeled via diffeomorphisms governed by elements in a reproducing kernel Hilbert space that satisfy minimal-energy constraints. Subject to the constraint, the deformation is determined by the key quantity "initial momenta," which are vector fields that represent the initial directions and magnitudes to "drag" each point on the baseline curve toward the target curve.

The variability in longitudinal changes of PA within a cohort of subjects is modeled through the variability of the deformations. Specifically, we adopt the functional principal component analysis (fPCA) to model the variability in subject-specific initial momenta. We focus on top principal components (PC) that are capable of explaining most of the variability in the cohort and examine the corresponding modes of variation in longitudinal changes in PA. In studying the longitudinal changes in PA of the subjects, of interest are the projection coefficients on the PCs for each subject, which characterize the composition of difference modes of changes in PA for each subject.

In modeling the relations between changes in PA and health outcomes/interventions, subjects' projection coefficients on PCs are used as the proxy for longitudinal changes in PA and are linked to the outcomes/interventions. We apply the proposed model and method to data from two clinical trials: RfH and MENU. For the RfH we use a linear model to examine the effect of the lifestyle intervention on projections on the top PCs while adjusting for relevant covariates selected by Lasso regularization. PCs characterizing different modes of changes in PA diurnal patterns, including an overall boosted PA throughout the day and change of PA levels in midday and evening, are discovered to be significantly associated with the intervention.

For the MENU study, we use a regression model to study the effect of different modes of changes in PA, as characterized by the top PCs, on the change in BMI while adjusting

for relevant covariates selected by Lasso regularization. A significant boost of morning PA levels is found to be significantly associated with decrease in BMI. A redistribution of PA with increased PA around noon and 3 p.m. is also found to be beneficial for weight loss. Moreover, findings of PA diurnal patterns are different in the MENU and RfH studies, indicating the importance of understanding detailed characteristics of PA in different study cohorts.

Models considered here can be readily generalized to study a variety of problems involving longitudinal changes in PA. For example, the effect of PA can be examined in a mediation analysis where the PC projections are potential mediators on the pathway from diet/lifestyle intervention to health outcomes. The PCs can also be studied in machine learning models to predict health outcomes.

Comparing to existing methods, the proposed model and method reveal important information from the minute-level activity tracker measured PA data and enable discovery of previously overlooked modes of changes in PA diurnal patterns. In particular, boost of activity in certain time windows and/or shifts of active hours to certain periods are found to be associated with lifestyle intervention and can be more effective in facilitating weight loss. Subjects with certain characteristics may have higher tendencies to experience changes in PA in certain modes. Such information is valuable in advising and treating patients and in designing individualized intervention programs and guidelines according to subject's personal medical history and conditions.

We would also like to emphasize the difference in the focus between the proposed approach and the time-warping/registration methods including Wrobel et al. (2019). In the latter it is assumed there exists an intrinsic time that is possibly different from the chronological time that indexes the diurnal patterns in physical activity. The time-warping approach aims to align the curves, usually repeatedly measured in the same period, based on the unobserved intrinsic time to reduce variability and to clarify underlying patterns. The focus of our study is to understand longitudinal changes in PA diurnal patterns, as indexed by the chronological time. In particular, shifts in PA patterns, including active hours, are under examination as a main quantity of interest.

The comparison between the two-step approach of separating temporal/phase and vertical/magnitude changes and the unified approach of modeling both changes simultaneously can also lead to many intriguing discussions. One important discussion is around the identifiability of the models. In a two-step approach, different choices of warping functions, used in the first step of phase registration, will lead to heterogeneity in the vertical change of PA magnitude in the second step and hence difference in the results. Essentially, there is an identifiability issue when separating the variability in the temporal domain and in the magnitudes. A relevant discussion can be found in Marron et al. (2015) and references therein, where the authors discuss the information gained from the unified approach comparing to the two-step approach. Some other interesting questions around this comparison include whether the two steps should be interchangeable in the two-step approach and the choice of warping based on parametric models and nonparametric models.

Due to the page limit, these questions are better suited for a thorough examination in future studies.

One of the advantages of directly modeling the change in PA via the diffeomorphism is that subjects' heterogeneity in each period's PA can be partially eliminated in estimating the subject-specific deformations between the PA curves in different periods. By directly modeling the change in PA, we are able to focus on the analysis of variability in the change with less interference from the heterogeneity in each individual period.

Potential sleep/wake-up time change is accounted for in the model of the diffeomorphisms. For example, if there is a change for a subject in the sleep and wake-up time between the baseline and the follow-up period, it will be reflected by a temporal shift of the earliest active time in the PA curve. In fact, the proposed model is versatile in characterizing not only changes in the active hours where PA happened but also changes in the transition between nonactive and active hours. If seasonality effect is of concern, adjustments can be achieved by a thorough exploratory analysis and preprocessing of the data. Adjustments can also be applied to the statistical inference involving the estimated diffeomorphisms.

In future studies we will explore details in modeling PA records collected from more than two visits. In addition, we are interested in further methodology development for statistical inference on changes in PA characterized by deformations of PA curves. For example, functional regression approaches (Goldsmith et al. (2011), Reiss and Ogden (2007), Crainiceanu, Staicu and Di (2009)) can be utilized to study effects of covariates on longitudinal changes in PA and/or effect of PA changes on health outcomes. In this paper we provide some first results of the MENU study with functional regression models in which the longitudinal change in PA, represented by the deformation momenta, is a functional covariate of interest. The deformation momenta on the x and y coordinates are modeled as separate covariates. In future studies we aim to explore modeling the deformation momenta as two-dimensional random fields and study theoretical properties of the model.

There can be multiple directions following this line of research based on the modeling of longitudinal changes in PA with diffeomorphisms. One of such directions is clustering of subjects in a study based on the modes of changes in PA patterns, where the clustering can be conducted based on similarity metrics constructed for the diffeomorphisms. Another possible direction is to study potential causal relationships between changes in PA and health outcomes and interventions including potential mediation effects. Alternative time-warping approaches to align PA curves incorporating changes in the PA magnitudes can also be developed based on this model.

Finally, example code, including estimation of deformation between PA curves in different periods, estimation of correlative models of PA, and health outcomes/interventions as well as simulation studies, has been published on the GitHub repository (https://github.com/jingjingz/PA_change) to share with global researchers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments.

Funding. DDS, SH, CLR, MJ and LN were partially supported by funding from the National Cancer Institute (U54 CA155435-01); JB, AC, MJ, JZ, and LN were partially supported by funding from the National Institute of Diabetes and Digestive and Kidney Disease (R01DK114945); JB, AC, CD, DDS, SH, and LN were partially supported by funding from the National Institute of Aging (PO1AG052352).

REFERENCES

Adamo KB, Prince SA, Tricco AC, Connor Gorber S and Tremblay M (2009). A comparison of indirect versus direct measures for assessing physical activity in the pediatric population: A systematic review. Int. J. Pediatr. Obes 4 2–27. [PubMed: 18720173]

Ainsworth B, Cahalin L, Buman M and Ross R (2014). The current state of physical activity assessment tools. Prog. Cardiovasc. Dis. 57 387–395. [PubMed: 25446555]

Amagasa S, Machida M, Fukushima N, Kikuchi H, Takamiya T, Odagiri Y and Inoue S (2018). Is objectively measured light-intensity physical activity associated with health outcomes after adjustment for moderate-to-vigorous physical activity in adults? A systematic review. Int. J. Behav. Nutr. Phys. Act 15 1–13. [PubMed: 29291739]

Anirudh R, Turaga P, Su J and Srivastava A (2016). Elastic functional coding of Riemannian trajectories. IEEE Trans. Pattern Anal. Mach. Intell 39 922–936. [PubMed: 28113699]

Bassett DR (2012). Device-based monitoring in physical activity and public health research. Physiol. Meas 33 1769–1783.

Beg MF, Miller MI, Trouve A and Younes L (2005). Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vis 61 139–157.

Bellettiere J, Healy GN, LaMonte MJ, Kerr J, Evenson KR, Rillamas-Sun E, Di C, Buchner DM, Hovell MF et al. (2019a). Sedentary behavior and prevalent diabetes in 6,166 older women: The objective physical activity and cardiovascular health study. J. Gerontol, Ser. A, Biol. Sci. Med. Sci 74 387–395. [PubMed: 29726906]

Bellettiere J, Lamonte MJ, Evenson KR, Rillamas-Sun E, Kerr J, Lee IM, Di C, Rosenberg DE, Stefanick ML et al. (2019b). Sedentary behavior and cardiovascular disease in older women: The opach study. Circulation 139 1036–1046. [PubMed: 31031411]

Bellettiere J, Lamonte MJ, Unkart J, Liles S, Laddu-Patel D, Manson JE, Banack H, Seguin-Fowler R, Chavez P et al. (2020). Short physical performance battery and incident cardiovascular events among older women. J. Amer. Heart Assoc 9 e016845. 10.1161/JAHA.120.016845 [PubMed: 32662311]

Charlier B, Charon N and Trouvé A (2017). The Fshape framework for the variability analysis of functional shapes. Found. Comput. Math 17 287–357. MR3627451 10.1007/s10208-015-9288-2

Chastin SFM, De Craemer M, De Cocker K, Powell L, Van Cauwenberg J, Dall P, Hamer M and Stamatakis E (2019). How does light-intensity physical activity associate with adult cardiometabolic health and mortality? Systematic review with meta-analysis of experimental and observational studies. Br. J. Sports Med 53 370–376. [PubMed: 29695511]

Choi H, Wang Q, Toledo M, Turaga P, Buman M and Srivastava A (2018). Temporal alignment improves feature quality: An experiment on activity recognition with accelerometer data. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops 349–357.

Colley RC, Garriguet D, Janssen I, Craig CL, Clarke J and Tremblay MS (2011). Physical activity of Canadian adults: Accelerometer results from the 2007 to 2009 Canadian health measures survey. Health Rep. 22 7–14.

Crainiceanu CM, Staicu A-M and Di C-Z (2009). Generalized multilevel functional regression. J. Amer. Statist. Assoc 104 1550–1561. MR2750578 10.1198/jasa.2009.tm08564

Di C-Z, Crainiceanu CM, Caffo BS and Punjabi NM (2009). Multilevel functional principal component analysis. Ann. Appl. Stat 3 458–488. MR2668715 10.1214/08-AOAS206 [PubMed: 20221415]

Dyrstad SM, Hansen BH, Holme IM and Anderssen SA (2014). Comparison of self-reported versus accelerometer-measured physical activity. Med. Sci. Sports Exerc 46 99–106. [PubMed: 23793232]

Dziak J and Shiyko M (2021). funreg: Functional regression for irregularly timed data. R package version 1.2.2.

Ekelund U, Tarp J, Steene-Johannessen J, Hansen BH, Jefferis B, Fagerland MW, Whincup P, Diaz KM, Hooker SP et al. (2019). Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: Systematic review and harmonised meta-analysis. BMJ 366 l4570. [PubMed: 31434697]

Füzéki E, Engeroff T and Banzer W (2017). Health benefits of light-intensity physical activity: A systematic review of accelerometer data of the national health and nutrition examination survey (NHANES). Sports Med. 47 1769–1793. [PubMed: 28393328]

Gajardo A, Carroll C, Chen Y, Dai X, Fan J, Hadjipantelis PZ, Han K, Ji H, Mueller H-G et al. (2021). Fdapace: Functional data analysis and empirical dynamics. R package version 0.5.7.

Glass NL, Bellettiere J, Jain P, LaMonte MJ, LaCroix AZ and Women's Health Initiative (2021). Evaluation of light physical activity measured by accelerometry and mobility disability during a 6-year follow-up in older women. JAMA Netw. Open 4 e210005. 10.1001/jamanetworkopen.2021.0005 [PubMed: 33620446]

Goldsmith J, Bobb J, Crainiceanu CM, Caffo B and Reich D (2011). Penalized functional regression. J. Comput. Graph. Statist 20 830–851. MR2878950 10.1198/jcgs.2010.10007

Goldsmith J, Liu X, Jacobson JS and Rundle A (2016). New insights into activity patterns in children, found using functional data analyses. Med. Sci. Sports Exerc 48 1723–1729. [PubMed: 27183122]

Goldsmith J, Zipunnikov V and Schrack J (2015). Generalized multilevel function-on-scalar regression and principal component analysis. Biometrics 71 344–353. MR3366239 10.1111/biom.12278 [PubMed: 25620473]

Greven S, Crainiceanu C, Caffo B and Reich D (2010). Longitudinal functional principal component analysis. Electron. J. Stat 4 1022–1054. MR2727452 10.1214/10-EJS575 [PubMed: 21743825]

Greven S, Crainiceanu C, Caffo B and Reich D (2011). Longitudinal functional principal component analysis. In Recent Advances in Functional Data Analysis and Related Topics. Contrib. Statist 149–154. Physica-Verlag/Springer, Heidelberg. MR2815575 10.1007/978-3-7908-2736-1_23

Hernandez M, Bossa MN and Olmos S (2009). Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows. Int. J. Comput. Vis 85 291–306.

Khan WAA, Jackson ML, Kennedy GA and Conduit R (2021). A field investigation of the relationship between rotating shifts, sleep, mental health and physical activity of Australian paramedics. Sci. Rep 11 1–11. [PubMed: 33414495]

Kurtek S. (2017). A geometric approach to pairwise Bayesian alignment of functional data using importance sampling. Electron. J. Stat 11 502–531. MR3619315 10.1214/17-EJS1243

LaCroix AZ, Bellettiere J, Rillamas-Sun E, Di C, Evenson KR, Lewis CE, Buchner DM, Stefanick ML, Lee I-M et al. (2019). Association of light physical activity measured by accelerometry and incidence of coronary heart disease and cardiovascular disease in older women. JAMA Netw. Open 2 e190419. [PubMed: 30874775]

LaMonte MJ, Buchner DM, Rillamas-Sun E, Di C, Evenson KR, Bellettiere J, Lewis CE, Lee I-M, Tinker LF et al. (2018). Accelerometer-measured physical activity and mortality in women aged 63 to 99. J. Amer. Geriatr. Soc 66 886–894. [PubMed: 29143320]

Le T, Flatt SW, Natarajan L, Pakiz B, Quintana EL, Heath DD, Rana BK and Rock CL (2016). Effects of diet composition and insulin resistance status on plasma lipid levels in a weight loss intervention in women. J. Amer. Heart Assoc

Li H, Kozey Keadle S, Staudenmayer J, Assaad H, Huang JZ and Carroll RJ (2015). Methods to assess an exercise intervention trial based on 3-level functional data. Biostatistics 16 754–771. MR3449841 10.1093/biostatistics/kxv015 [PubMed: 25987650]

Li H, Staudenmayer J and Carroll RJ (2014). Hierarchical functional data with mixed continuous and binary measurements. Biometrics 70 802–811. MR3295741 10.1111/biom.12211 [PubMed: 25134936]

Loprinzi PD (2016). Light-intensity physical activity and all-cause mortality. Am. J. Health Promot 31 340–342. [PubMed: 26730555]

Marron JS, Ramsay JO, Sangalli LM and Srivastava A (2015). Functional data analysis of amplitude and phase variation. Statist. Sci 30 468–484. MR3432837 10.1214/15-STS524

Matthews CE, Ockene IS, Freedson PS, Rosal MC, Merriam PA and Hebert JR (2002). Moderate to vigorous physical activity and risk of upper-respiratory tract infection. Med. Sci. Sports Exerc 34 1242–1248. [PubMed: 12165677]

Migueles JH, Cadenas-Sanchez C, Ekelund U, Nyström CD, Mora-Gonzalez J, Löf M, Labayen I, Ruiz JR and Ortega FB (2017). Accelerometer data collection and processing criteria to assess physical activity and other outcomes: A systematic review and practical considerations. Sports Med. 47 1821–1845. [PubMed: 28303543]

Montaruli A, Galasso L, Caumo A, Cè E, Pesenti C, Roveda E and Esposito F (2017). The circadian typology: The role of physical activity and melatonin. Sport Sci. Health 13 469–476.

Nader PR, Bradley RH, Houts RM, McRitchie SL and O'Brien M (2008). Moderate-to-vigorous physical activity from ages 9 to 15 years. JAMA 300 295–305. [PubMed: 18632544]

Parada H, McDonald E, Bellettiere J, Evenson KR, LaMonte MJ and LaCroix AZ (2020). Associations of accelerometer-measured physical activity and physical activity-related cancer incidence in older women: Results from the WHIOPACH study. Br. J. Cancer 122 1409–1416. [PubMed: 32139875]

Patterson RE, Marinac CR, Natarajan L, Hartman SJ, Cadmus-Bertram L, Flatt SW, Li H, Parker B, Oratowski-Coleman J et al. (2016). Recruitment strategies, design, and participant characteristics in a trial of weight-loss and metformin in breast cancer survivors. Contemp. Clin. Trials 47 64–71. [PubMed: 26706665]

Patterson RE, Marinac CR, Sears DD, Kerr J, Hartman SJ, Cadmus-Bertram L, Villaseñor A, Flatt SW, Godbole S et al. (2018). The effects of metformin and weight loss on biomarkers associated with breast cancer outcomes. J. Natl. Cancer Inst 110 1239–1247. [PubMed: 29788487]

Prince SA, Adamo KB, Hamel M, Hardt J, Connor Gorber S and Tremblay M (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: A systematic review. Int. J. Behav. Nutr. Phys. Act 5 56–24. [PubMed: 18990237]

Ramakrishnan R, Doherty A, Smith-Byrne K, Rahimi K, Bennett D, Woodward M, Walmsley R and Dwyer T (2021). Accelerometer measured physical activity and the incidence of cardiovascular disease: Evidence from the uk biobank cohort study. PLoS Med. 18.

Reiss PT and Ogden RT (2007). Functional principal component regression and functional partial least squares. J. Amer. Statist. Assoc 102 984–996. MR2411660 10.1198/016214507000000527

Reuter C, Bellettiere J, Liles S, Di C, Sears DD, LaMonte MJ, Stefanick ML, LaCroix AZ and Natarajan L (2020). Diurnal patterns of sedentary behavior and changes in physical function over time among older women: A prospective cohort study. 1–11.

Rock CL, Flatt SW, Pakiz B, Quintana EL, Heath DD, Rana BK and Natarajan L (2016). Effects of diet composition on weight loss, metabolic factors and biomarkers in a 1-year weight loss intervention in obese women examined by baseline insulin resistance status. Metab. Clin. Exper 65 1605–1613. [PubMed: 27733248]

Shou H, Zipunnikov V, Crainiceanu CM and Greven S (2015). Structured functional principal component analysis. Biometrics 71 247–257. MR3335369 10.1111/biom.12236 [PubMed: 25327216]

Srivastava A, Wu W, Kurtek S, Klassen E and Marron JS (2011). Registration of functional data using fisher-rao metric. arXiv preprint arXiv:1103.3817.

Stamatakis E, Gale J, Bauman A, Ekelund U, Hamer M and Ding D (2019). Sitting time, physical activity, and risk of mortality in adults. J. Am. Coll. Cardiol 73 2062–2072. 10.1016/j.jacc.2019.02.031 [PubMed: 31023430]

R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Vaillant M, Miller MI, Younes L and Trouvé A (2004). Statistics on diffeomorphisms via tangent space representations. NeuroImage 23 S161–S169. [PubMed: 15501085]

Walker RL, Greenwood-Hickman MA, Bellettiere J, LaCroix AZ, Wing D, Higgins M, Richmire KR, Larson EB, Crane PK et al. (2021). Associations between physical function and device-based

measures of physical activity and sedentary behavior patterns in older adults: Moving beyond moderate-to-vigorous intensity physical activity. BMC Geriatr. 21.

Wrobel J, Bauer A, McDonnell E and Goldsmith J (2022). registr: Curve registration for exponential family functional data. R package version 2.1.0.

Wrobel J, Zipunnikov V, Schrack J and Goldsmith J (2019). Registration for exponential family functional data. Biometrics 75 48–57. MR3953706 10.1111/biom.12963 [PubMed: 30129091]

Xiao L, Huang L, Schrack JA, Ferrucci L, Zipunnikov V and Crainiceanu CM (2015). Quantifying the lifetime circadian rhythm of physical activity: A covariate-dependent functional approach. Biostatistics 16 352–367. MR3365433 10.1093/biostatistics/kxu045 [PubMed: 25361695]

Xu SY, Nelson S, Kerr J, Godbole S, Johnson E, Patterson RE, Rock CL, Sears DD, Abramson I et al. (2019). Modeling temporal variation in physical activity using functional principal components analysis. Stat. Biosci 11 403–421.

Zou J, Lin T, Di C, Bellettiere J, Jankowska MM, Hartman SJ, Sears DD, LaCroix AZ, Rock CL and Natarajan L (2023). Supplement to “A Riemann manifold model framework for longitudinal changes in physical activity patterns”. 10.1214/23-AOAS1758SUPP

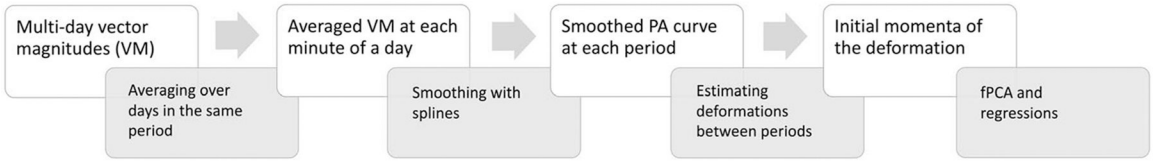

Fig. 1.
Steps in the estimation pipeline.

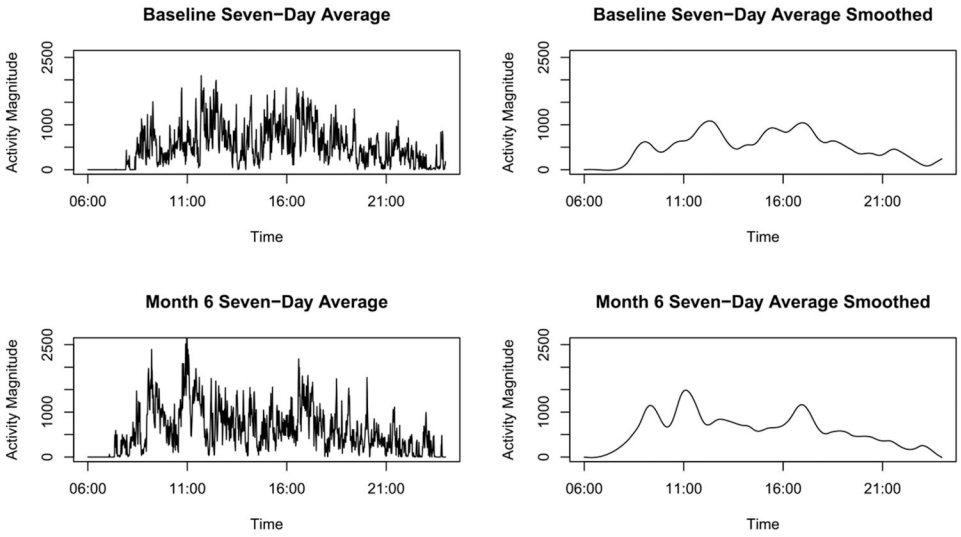

Fig. 2.

Example of PA curves at baseline and month-6 before and after smoothing for one subject in the RfH study.

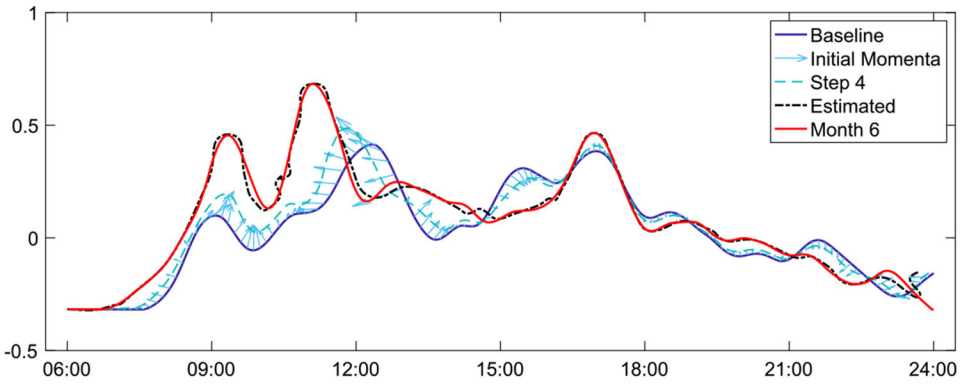

Fig. 3.

RfH study: *Estimated deformations of one example subject. Baseline PA curve* (*solid line*) *on which estimated initial momenta at each control point are shown with arrows, middle step* ($\tau = 4 / 11$: *the fourth step in* 11 *total discretized steps of* τ. *Details in* Section 6.2 *of deformation* (*dashed lines*) *and estimated* (*dashed*) *overlapping mostly with actual month*-6 *curves* (*solid*). *Y-axis*: *Centered and scaled PA magnitude. X-axis*: *Time of a day*.

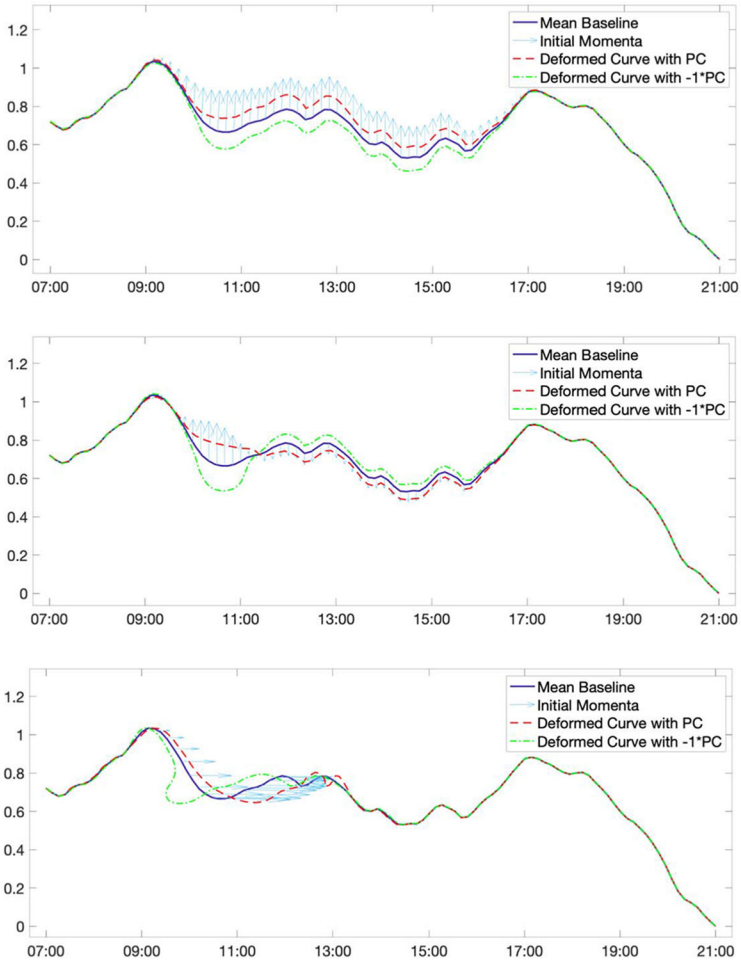

Fig. 4.

From top to bottom: *Actual initial momenta* (*arrows*) *of PC* 1, 2, *and* 3 *and deformations when applying the PCs to the mean baseline curve* (*solid*) *to deform into the follow-up curves. Dashed lines are deformed curves with* $1 \times PC$ *and* $-1 \times PC$.

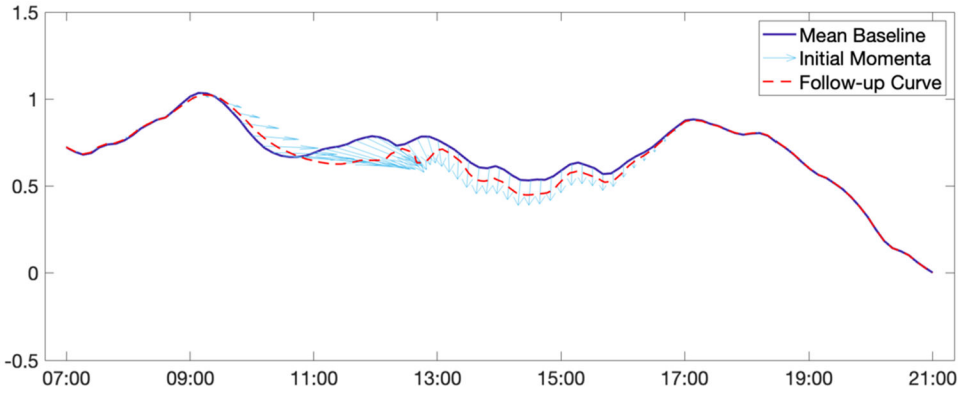

Fig. 5.
Baseline and follow-up curves of an exemplar subject.

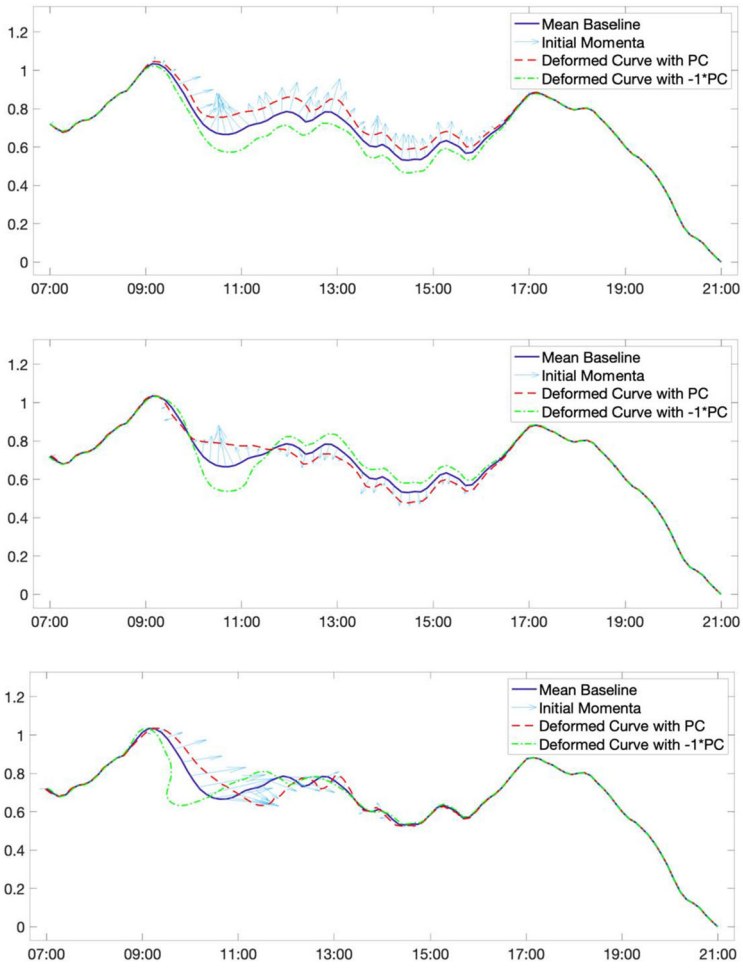

Fig. 6.

From top to bottom: *Estimated initial momenta with proposed approach* (*arrows*) *of PC* 1, 2, *and* 3 *and associated deformations when applying the momenta to the mean baseline curve* (*solid*) *to deform into the follow-up curves* (*dashed lines for* 1 *and* $-1 \times$ *PC*).

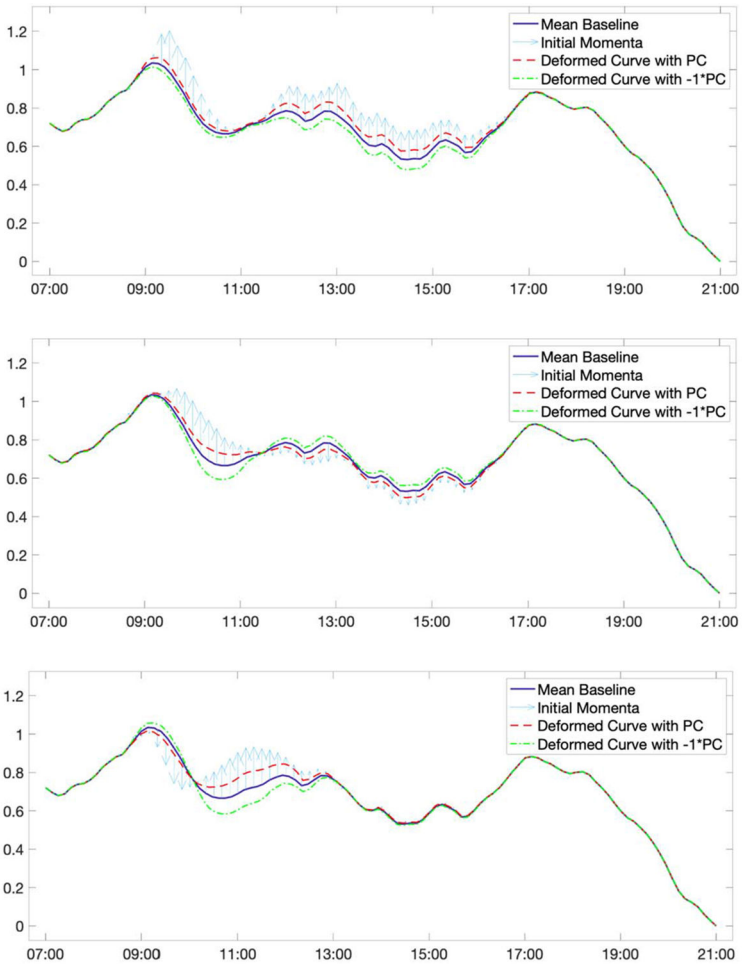

Fig. 7.

From top to bottom: *Estimated initial momenta* (*arrows*) *of PCs* 1, 2 *and* 3 *extracted from vertical differences between baseline and follow-up PA curves, and deformations when applying the PCs to the mean baseline curve* (*solid*) *to deform into the follow-up curves* (*dashed lines for* 1 *and* $-1 \times PC$).

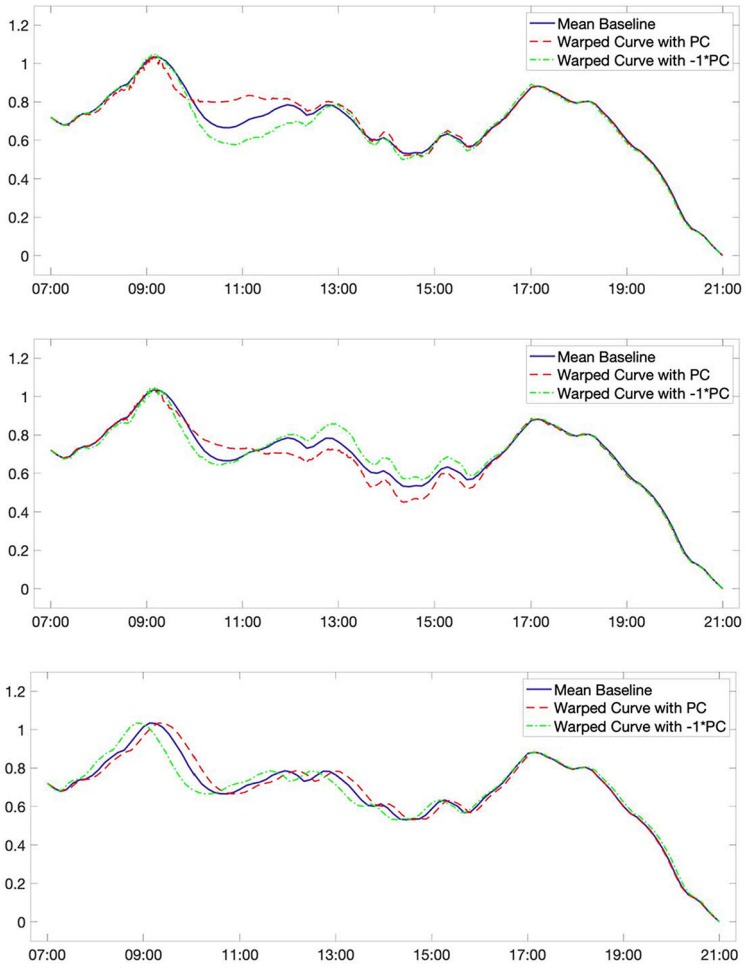

Fig. 8.

Top and middle: *Estimated follow-up curves associated with PC* 1 *and* 2 *extracted from difference in PA magnitude between the follow-up and time-warped baseline curve. Bottom*: *Follow-up curve estimated with PC* 1 *of time-warping of the mean baseline curve. Solid*: *Baseline curve. Dashed lines*: *Estimated follow-up curves associated with* 1 *and* (−1) *times the PCs, scaled by the standard deviation*.

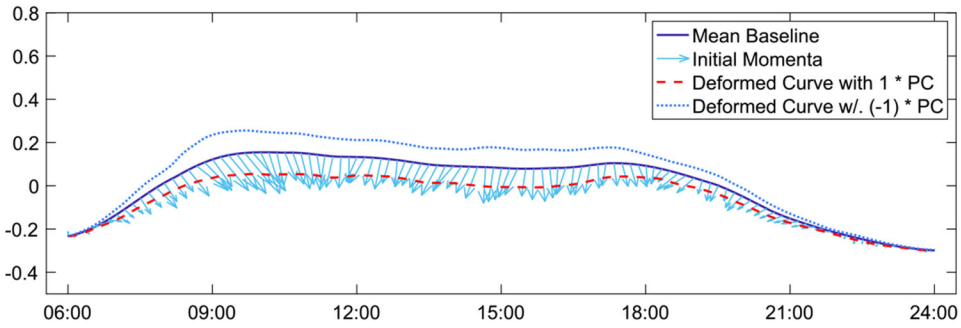

Fig. 9.

RfH study: *Estimated deformations of PC* 1. *Baseline* (*solid*), *estimated initial momenta* (*arrows on top of the baseline curve*) *and estimated month*-6 *curves* (*dashed*) *by applying the deformations corresponding to* +1 *and* −1 *times the initial momenta of the PC*.

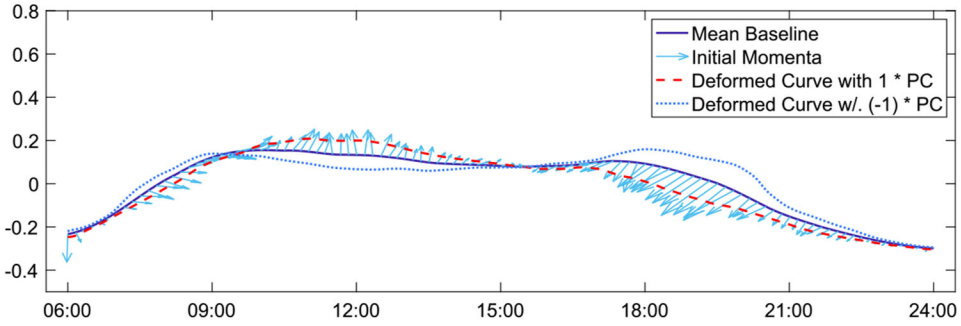

Fig. 10.

RfH study: *Estimated deformations of PC* 3.

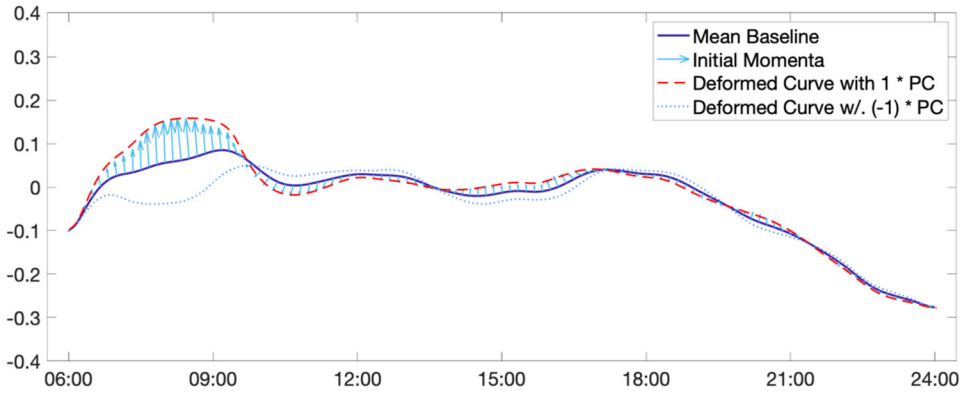

Fig. 11.

MENU Study: *Estimated deformations of PC* 1.

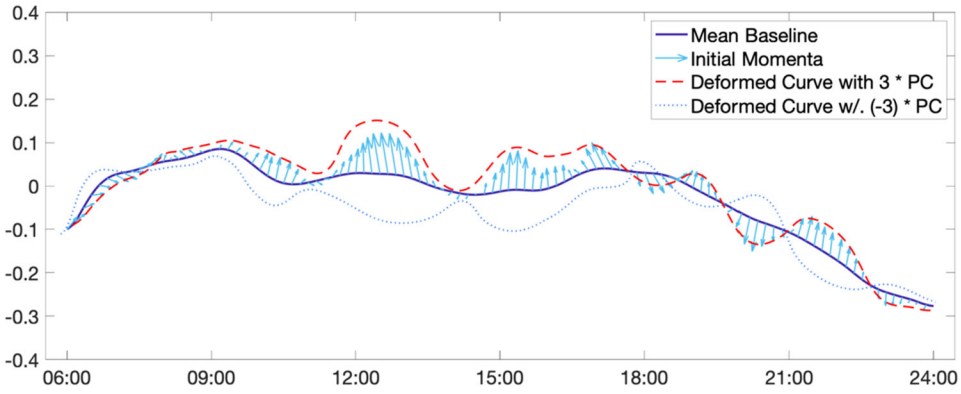

Fig. 12.

MENU Study: *Estimated deformations of PC* 10.

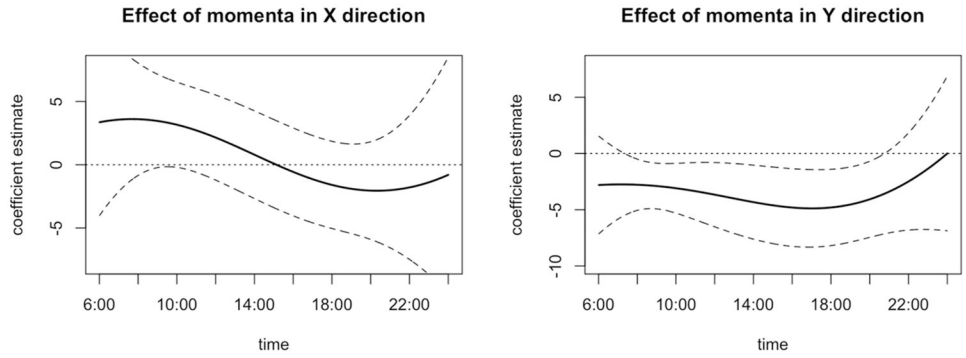

Fig. 13.
Estimates of functional coefficients adjusting for covariates.

Table 1

RfH study: *Coefficient estimates* (*with standard errors*) *and p-values of covariates' effects on PC* 1 *projection scores*

Covariate	Coef Estimate (SE)	p-value
(Intercept)	−3.617 (0.920)	<0.001 ***
Lifestyle Intervention	−0.265 (0.108)	0.015 *
AGE	0.027 (0.027)	0.312
Age at diagnosis	0.005 (0.026)	0.843
SED	−0.001 (0.001)	0.504
LPA100	0.004 (0.001)	<0.001 ***
MVPA1952	0.008 (0.003)	0.012 *
Leptin	0.005 (0.002)	0.036 *
Baseline Weight	0.005 (0.004)	0.227
Insomnia Yes	−0.110 (0.118)	0.352
High Cholesterol Yes	0.198 (0.108)	0.066 .
Diabetes Yes	− 1.012 (0.355)	0.005 **
Aspirin Yes	−0.373 (0.127)	0.003 **
Education Graduated from high school or G.E.D.	−0.408 (0.222)	0.067 .
Marital Living with a partner in a marriage like relationship	−0.408 (0.282)	0.149
Employment Disabled and/or retired because of health	0.449 (0.220)	0.043 *

Table 2

RfH study: *Coefficient estimates* (*with standard errors*) *and p-values of covariates' effects on PC* 3 *projection scores*

Covariate	Coef Estimate (SE)	p-value
(Intercept)	−1.076 (0.642)	0.095 .
Lifestyle Intervention	−0.195 (0.112)	0.082 .
Wear time	0.002 (0.001)	0.033 *
Smoke Status Former	−0.365 (0.112)	0.001 **
Employment Disabled and/or retired because of health	−0.460 (0.232)	0.048 *

Table 3

MENU study: *Coefficient estimates* (*with standard errors*) *and p-values on* $\mathrm{BMI}_{month\ 12} - \mathrm{BMI}_{\mathrm{baseline}}$

Covariate	Coef Estimate (SE)	p-value
Intercept	−0.060 (0.825)	0.942
Diet Lower Fat	−1.123 (0.479)	0.020 *
Diet Walnut-Rich	−0.772 (0.498)	0.123
Black	0.608 (0.842)	0.471
Asian	−1.043 (1.478)	0.481
Pacific Islander	−2.112 (1.820)	0.248
Native American	−2.380 (2.610)	0.363
Mixed Race	−0.978 (1.470)	0.507
Other Race	3.888 (2.542)	0.128
Single-never married	−1.054 (0.578)	0.070 .
Widowed	−2.779 (1.301)	0.034 *
Divorced	−0.187 (0.557)	0.738
Separated	−0.977 (2.576)	0.705
High blood pressure	0.229 (0.612)	0.709
High blood pressure and High cholesterol	0.457 (0.620)	0.462
High blood pressure and other	0.845 (1.806)	0.640
High cholesterol	1.841 (0.629)	0.004 **
Other	0.943 (1.178)	0.424
Pain SF-36 subscale	−0.018 (0.009)	0.051 .
Follicle stimulating hormone (FSH)	−0.018 (0.006)	0.003 **
PC1	−0.507 (0.195)	0.010 *
PC10	−0.428 (0.192)	0.027 *
PC20	−0.404 (0.200)	0.045 *
PC25	−0.391 (0.199)	0.051 .