UCLA
UCLA Previously Published Works

Title

BAYESIAN HIERARCHICAL MODELING AND ANALYSIS FOR ACTIGRAPH DATA FROM WEARABLE DEVICES.

Permalink

https://escholarship.org/uc/item/7v56v12g

Journal

The Annals of Applied Statistics, 17(4)

ISSN

1932-6157

Authors

Di Loro, Pierfrancesco Alaimo

Mingione, Marco

Lipsitt, Jonah

et al.

Publication Date

2023-12-01

DOI

10.1214/23-aoas1742

Copyright Information

Peer reviewed

BAYESIAN HIERARCHICAL MODELING AND ANALYSIS FOR ACTIGRAPH DATA FROM WEARABLE DEVICES

BY PIERFRANCESCO ALAIMO DI LORO [1,*], MARCO MINGIONE [2,†], JONAH LIPSITT [3,‡], CHRISTINA M. BATTEATE [4,§], MICHAEL JERRETT [3] AND SUDIPTO BANERJEE [5]

[1]*Department GEPLI, LUMSA, *p.alaimodiloro@lumsa.it*

[2]*Department of Political Sciences, Roma Tre University, †marco.mingione@uniroma3.it*

[3]*Department of Environmental Health Sciences, University of California, Los Angeles, ‡jonahlipsitt@gmail.com; mjerrett@ucla.edu*

[4]*Center of Occupational and Environmental Health, University of California, Los Angeles, §cbatteate@ucla.edu*

[5]*Department of Biostatistics, University of California, Los Angeles, sudipto@ucla.edu*

The majority of Americans fail to achieve recommended levels of physical activity, which leads to numerous preventable health problems such as diabetes, hypertension, and heart diseases. This has generated substantial interest in monitoring human activity to gear interventions toward environmental features that may relate to higher physical activity. Wearable devices, such as wrist-worn sensors that monitor gross motor activity (actigraph units) continuously record the activity levels of a subject, producing massive amounts of high-resolution measurements. Analyzing actigraph data needs to account for spatial and temporal information on trajectories or paths traversed by subjects wearing such devices. Inferential objectives include estimating a subject's physical activity levels along a given trajectory; identifying trajectories that are more likely to produce higher levels of physical activity for a given subject; and predicting expected levels of physical activity in any proposed new trajectory for a given set of health attributes. Here, we devise a Bayesian hierarchical modeling framework for spatial-temporal actigraphy data to deliver fully model-based inference on trajectories while accounting for subject-level health attributes and spatial-temporal dependencies. We undertake a comprehensive analysis of an original dataset from the Physical Activity through Sustainable Transport Approaches in Los Angeles (PASTA-LA) study to ascertain spatial zones and trajectories exhibiting significantly higher levels of physical activity while accounting for various sources of heterogeneity.

1. Introduction. Promoting a healthy lifestyle continues to stoke substantial research activities in public health. The "Physical Activity Guidelines for Americans" (2nd edition) suggests that most individuals, depending on age and body composition, receive 150-300 minutes of moderate to vigorous physical activity (MVPA) weekly (Piercy et al., 2018). In general, the scientific community agrees that regular physical activity (PA) can have immediate and long-term health benefits (Reiner et al., 2013; Bull et al., 2020). Despite these well-known benefits, most Americans fail to meet recommended requirements (Piercy et al., 2018). Specifically, only 1 in 5 high-school adolescents and 1 in 4 adults meet recommended levels of physical activity. Given the well-established relationships between lack of PA and several obesity-related chronic conditions such as heart disease, type 2 diabetes, and can-

cer, as well as many physical and mental health benefits, an urgent need exists to improve monitoring of PA and to establish public health programs that promote more PA[1].

Technologies for monitoring spatial energetics (James et al., 2016; Drewnowski et al., 2020) and promoting physical activity continue to emerge. Among others, *actigraphy* broadly refers to the monitoring of human rest and activity cycles using wearable devices. Actigraphy data are gathered directly from wearable sensors or indirectly through smart-phone mobile applications and record repeated measurements at a very high resolutions. In particular, accelerometers are motion sensors that measure acceleration along different axes and are able to collect large amounts of data (Plasqui and Westerterp, 2007; Sikka et al., 2019). They are increasingly conspicuous because of their affordability, accuracy, and availability in smart-phones, smart-watches and other wearable devices. Many devices also include Global Positioning System (GPS) sensors that reference measurements with location tracking along trajectories, or paths, traversed by the subject. Collected data can be quickly downloaded and promptly analyzed to obtain insights into their pattern and structure.

We pursue a comprehensive analysis of an original actigraphy data set from the Physical Activity through Sustainable Transport Approaches in Los Angeles (PASTA-LA) study. Actigraphy and GPS data analysis customarily involve *idle* records that occur if a charged device does not detect acceleration over a specified time interval (e.g., 10 seconds). While idle records may correspond to periods of a subject's inactivity, they can also arise from other factors including technical malfunctions or the subject not wearing the device. On the other hand, eliminating idle records does not exclude all inactive periods because the accelerometer still records minor movements from where it is worn while a subject may be mostly inactive. Attempting to account for idle records as representative of inactive periods are likely to confound assessments of a subject's activity levels with current technological capabilities. Hence, we do not consider idle records and focus on the following specific data analytic aims: (i) estimating a subject's physical activity levels along trajectories; (ii) identifying trajectories that are more likely to produce higher levels of physical activity for a given subject; and (iii) predicting expected levels of physical activity in any proposed new trajectory for a given set of health attributes. Researchers find actigraph tracking especially attractive as it allows for a better understanding of what behavioral and environmental factors influence population and individual health and, hence, aid in public health recommendation and policy.

Actigraphs generate data evolving over space and time, which suggests rich classes of space-time models for analysis (Gelfand et al., 2010; Cressie and Wikle, 2015). In particular, actigraph analysis presents some notable challenges (Kestens et al., 2017): the data sets are large, or even massive, as they are recorded at very high frequencies; they exhibit dependence along trajectories which should be accounted for both explanation and prediction (Ray et al., 2018; Bai et al., 2018). We argue against a customary spatial-temporal process over $\mathbb{R}^2 \times \mathbb{R}^+$ and propose disentangling spatial effects from temporal dependence along trajectories. The balance of the paper is organized as follows. Section 2 introduces the PASTA-LA data-set with insights into accelerometry data. The model for the temporal correlation is introduced in Section 3, while spatial effects are discussed in Section 3.4. An extensive simulation study validating our model is proposed in Section 3.5. Data analysis, model assessment and comparisons are presented in Section 4. Finally, we conclude with a discussion in Section 5.

2. Data. Our data set is compiled from the original **P**hysical **A**ctivity through **S**ustainable **T**ransport **A**pproaches in **L**os **A**ngeles (PASTA-LA) study conducted on a cohort of 460 individuals monitored between May 2017 and June 2018. Data were collected through different

---

[1]More details at https://www.cdc.gov/chronicdisease/resources/publications/factsheets/physical-activity.htm

sources: online questionnaires, a smartphone app named *MOVES*, a GPS device (Global-Sat DG-500), and a wearable actigraph unit (Actigraph GT3X+). Data collected through the *MOVES* app, whose reliability must still be verified and discussed, are not considered in this paper. While 460 is the sample size of the complete study, the GPS and actigraph devices were deployed only on a nested sample of 134 individuals due to cost considerations. We analyze data collected through these two devices that were supposed to be worn by the participants in the nested sample for two one-week periods (one in 2017 and one in 2018). Study protocols for safeguarding participant information received necessary institutional review board (IRB) approval. The data were stored on a secure computer and a redacted version was created for purposes of data sharing.

2.1. *Questionnaires.* The online questionnaires included two baseline and four follow-up surveys: one baseline and two follow-ups for each collection period of the actigraph and GPS data. Each survey consisted of responses pertaining to the participant's demographics and transportation habits. Here, we consider the *first baseline questionnaire*, which is the only one available for all the participants in the nested sample. Personal information and other socioeconomic factors have been encoded as follows for subsequent analysis:

- **Sex**: Female or Male;
- **Ethnicity**: Asian, Black/African/Caribbean, Latin-American, White, or Other (mixed multiple ethnic groups or prefer not to answer);
- **Age (years) class**: $(0, 18]$, $(18, 25]$, $(25, 34]$, $(34, 45]$, $(45, 70]$;
- **BMI** (kg/m^2) **class**[2]: underweight if BMI $\in (15, 18.5]$, normal if BMI $\in (18.5, 24.5]$, overweight if BMI $\in (24.5, 30]$, and obese if BMI > 30;
- **Yearly Income Level** (in thousands \$): $(0, 10]$, $(10, 25]$, $(25, 50]$, $(50, 75]$, $(75, 100]$, $(100, 150]$, $(150, +\infty]$, and Don't know/Prefer not to answer;
- **Educational attainment**: High-school diploma, College graduate, Associate degree, Graduate;

We filtered unreasonable values of the BMI, i.e. BMI< 10, which was observed just for one individual, leaving 133 out of 134 individuals in the nested sample. A user ID was assigned to each survey response data and a redacted master key was generated using all ID types for joining with other study data.

2.2. *Actigraph.* The Actigraph unit is an accelerometer roughly the same size and weight of the average wrist-watch. It can be worn on the wrist, hip, and thigh and measure the directional acceleration at a specified time frequency (up to 100 Hz). The Actigraph GT3X+ model used for the PASTA-LA study can detect accelerations measured in gravitational units (G) with a sensitivity of ± 3 milligravity (mG) in the three orthogonal planes (anteroposterior, mediolateral, and vertical). Data are stored in an internal memory and can be downloaded to other hardware for analysis through a proprietary software. The participants were asked to start wearing the accelerometer on their dominant wrist as soon as it was handed to them, as the devices could have been properly calibrated at that time. The study protocol demanded that participants wear the Actigraph unit at all times other than during bathing and sleeping (awake time was assumed approximately from 7am to 11pm). The sampling frequency has been set to 30 Hz and the *idle sleep* mode has been activated in order to save battery and memory. With this mode on, the device would go idle every-time it records no acceleration ($< \pm 40 mG$) for 10 consecutive seconds. The Actigraph GT3X+ grants access

---

[2]according to standard guidelines of the Center of Disease Control and Prevention https://www.cdc.gov/obesity/basics/adult-defining.html

to the a `.gt3x` file with the raw acceleration measurements. It can be loaded in `R` using the `read.gt3x` package and contains the raw accelerations at each timestamp. Such accelerations comprise the basic ingredients to get a proxy for body movement from an accelerometer (Mathie et al., 2003; Migueles et al., 2019; Bammann et al., 2021). There are substantial investigations into its statistical relationships with PA measures, such as *energy expenditure* measures (EE) (Crouter, Clowers and Bassett Jr, 2006; Freedson et al., 2012; Taraldsen et al., 2012) and the *Metabolic Equivalent of Task* (MET) (Lyden et al., 2014; Staudenmayer et al., 2015; Migueles et al., 2017; van Loo et al., 2018). Among various metrics, we take the instantaneous body vector Magnitude of Acceleration (MAG) as the primary endpoint of our analysis (van Hees et al., 2011; White et al., 2016; Doherty et al., 2017). Further discussion about the conversion of MAG into energy expenditure measures is reported in Section 7 of the Supplementary Material (Alaimo Di Loro et al., 2023).

We were able to retrieve the Actigraph raw data only on $K = 97$ out of the 133 original individuals. Let x, y and z be the dynamic acceleration of the body of the k-th individual. The point-wise MAG is defined as:

$$\text{MAG}_{kt} = \sqrt{x_{kt}^2 + y_{kt}^2 + z_{kt}^2}, \qquad k = 1, \ldots, K. \tag{1}$$

However, the raw accelerations recorded by the accelerometer must be appropriately processed to glean body movement (Doherty et al., 2017). Indeed, the raw acceleration recorded by each axle is the sum of both the *static* and the *dynamic* acceleration, but only the second is the effect of actual body movement. First, we remove *idle* records, i.e. all the occasions in which the accelerometer recorded zero acceleration for longer than 15 seconds and the device went *idle*. It is very unlikely that these idle records with zero acceleration coincide with a subject's inactive periods because the accelerometer still records positive, albeit small, magnitudes of accelerations over inactive periods due to movements in the wrists, hips and thighs. Idle records, on the other hand, are likely to arise from technical malfunctions or from a subject violating protocol and not wearing the device in the experimental time window.

Second, the raw accelerations recorded by single axles must be disentangled from unwanted static or non-static components: the effect of the earth's gravitational force and other external accelerations (e.g car, bus, elevators) at low frequencies, machine noise and vibrations at high frequencies. To address this issue, we adopted a Band-Pass Butterworth digital filter of order 4 with frequency window $(0.25, 10)$ to clean the signals from these long and short waves (Mathie et al., 2003). Indeed, most human activities result in signals with a frequency between 0.25Hz and 10Hz (Khusainov et al., 2013). An example of how the raw signal is modified through this process is provided in Figure 9 of the Supplementary Material (Alaimo Di Loro et al., 2023). We subsequently evaluate the point-wise MAG_{kt} using the filtered accelerations $(\widetilde{x}_{kt}, \widetilde{y}_{kt}, \widetilde{z}_{kt})$. However, the instantaneous MAG evaluated at the original 30Hz frequency is extremely erratic and the single value may not represent well the PA intensity of the participant at that time. For this reason, it is usually averaged over 5 to 10 second epochs to acquire a more suitable measure of PA intensity (Migueles et al., 2017; Doherty et al., 2017). Here, we perform a kernel smoothing of the 30Hz measurements in order to be representative of the single time-point, and get a 1 second time resolution. The resulting vector magnitude is

$$\widetilde{\text{MAG}}_{kt} = \sum_{j} k_b(t - t_j) \cdot \text{MAG}_{kt_j}, \tag{2}$$

where $k_b(\cdot)$ is a Gaussian kernel with bandwidth $b = 5$ seconds. This ensures that the impact of the neighboring points becomes negligible for $|t - t_j| > 10$ seconds (see Figure 10 of the Supplementary Material (Alaimo Di Loro et al., 2023) as an example). Finally, we removed

all the observations recorded outside of the pre-specified daily time-window, i.e. from 7am to 11pm.

2.3. *GPS.* The GPS device *GlobalSat DG-500* recorded the subject's location (latitude and longitude) roughly every 5 seconds, together with date, time, and speed (km/h, measured as distance over time through linear interpolation). This work restricts the attention to 93 out of 97 subjects living and working in the Westwood neighborhood of Los Angeles in order to avoid a geographical imbalance that could bias and invalidate the model estimates. This area hosts the university campus of UCLA and it includes the largest part of all the available observations. Westwood is a walk-friendly neighborhood with a lot of green areas, parks, and major roads with shops and amenities. People were free to move inside and between buildings (e.g. people at the gym, office, etc) and we are interested in quantifying their movement in all these settings.

However, GPS measurements can be affected by possible inaccuracies, especially around buildings, that may cause unreasonable jumps in a very small time-span. We note that most of these issues are already mitigated by an automatic filtering process of the GPS device, that would drop records for which the signal is not strong enough. Nevertheless, to further enhance the cleaning process of GPS measurements, we removed all data points for which the computed average speed between two subsequent points was larger than $100km/h$. We picked such a high threshold as we do not want to drop observations related to individuals standing or sitting in a bus or car.

2.4. *External covariates.* PA levels are not only affected by individual characteristics, but can be fostered by specific features of the surrounding area they are navigating. Therefore, we included three external covariates to account for some of the built-in environment features of the Westwood area. In particular, we used the following.

(i) The weighted overlay distance to parks (in km) with a spatial resolution of 23×23 which can be downloaded from https://egis2.lacounty.gov/arcgis. It represents a weighted distance of each point from officially recognized parks and it can be seen as a proxy of the *green area density* (see Figure 1a; darker shades indicate proximity to parks).

(ii) The Normalized Difference Vegetation Index (NDVI), which is available with a spatial resolution of 30×30, can be downloaded from https://earthexplorer.usgs.gov/ and provides a measure of the *greeness* of the patch itself (see Figure 1b; darker shades depict more greenness).

(iii) The slope (azimuth), with a spatial resolution of 23×23, is computed from the digital elevation model (DEM) downloaded from http://www.webgis.com/terr_pages/CA/dem1/losangeles.html. It represents the average angular inclination of the ground patch with respect to the horizon line (see Figure 1c; darker shades depict higher slopes).

While previous studies (e.g., Maddison et al., 2009) have reported on these variables affecting PA levels, they usually consider the environmental impact on the average PA level through a buffer around the home location of the participant, and not on its instantaneous PA level. With our current work, we want to discover and establish direct associations between PA levels of a subject and these spatially-indexed covariates along trajectories.
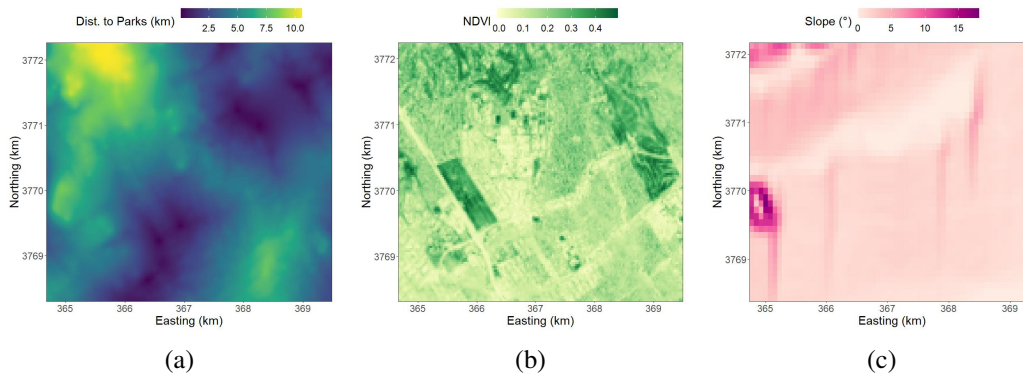
Fig 1: Spatial-varying covariates: (a) distance to parks, (b) NDVI and (c) Slope.

2.5. *Joining.* GPS and accelerometer data were all assigned a participant ID aligned with the questionnaires' master-key to facilitate joining across all ID types (including email) while redacting and encrypting user data. The first baseline questionnaire, Actigraph, and GPS were available for the aforementioned group of $K = 93$ individuals. Henceforth, we refer to this specific group of units. The joining of different data sources follows these steps.

1. Actigraph data are joined to the first baseline questionnaire using the individual master-key. The resulting data set includes the physical activity endpoint recorded by the Actigraph at the different timestamps and all the available individual information, but no spatial information.
2. GPS data are joined to the dataset obtained in 1) using the individual master-key and the timestamp. Following the processing in Section 2.2, the Actigraph data are available at the same time resolution of GPS data as long as the subject wore them simultaneously. Therefore, we decided to use the GPS as the leading table in the joining process. This avoids use of artificial data (e.g. interpolating GPS locations).
3. Spatial covariates are joined to the data set obtained in 2) through the *minimum distance* criteria, i.e. each location is assigned the value of the closest point on the grid for each spatial-varying covariate.

The temporal coverage is not balanced across individuals because (i) subjects move around Westwood in different segments of the overall time window; and (ii) some participants violate the study protocol. Indeed, not all the participants were available for both of the one-week surveillance periods in 2017 and 2018. In fact, only 58 out of the 93 participants have data for the first week only and missed the follow-up survey. In the end, we go from the least represented individuals having $\approx 5 \times 10^2$ observations (≈ 2 hours of data) to the most represented ones with more than $\approx 5 \times 10^5$ observations (≈ 14 days of data). Considering all the 5 second time segments between the first and last observed point of each individual in each day as the potential observation window, the proportion of missing measurements ranges between $\approx 31\%$ and $\approx 97\%$. The overall proportion of missing measurements in the entire database (based upon aggregated 5-second time segments for all individuals) is $\approx 83\%$. Figure 2 shows the number of observations for each individual $k = 1, \ldots, K$ in each hour $h = 7, \ldots, 22$, where the y-axis has been ordered in ascending order according to proportion of missing measurements for each subject. Overall, we can state that all hours are well-represented, but only few individuals have data for the whole daily time window. Summing up, all subsequent analysis will refer to the final data set consisting of $N \simeq 7 \times 10^5$ measurements across $K = 93$ individuals, scattered over Westwood (see Figure 3).
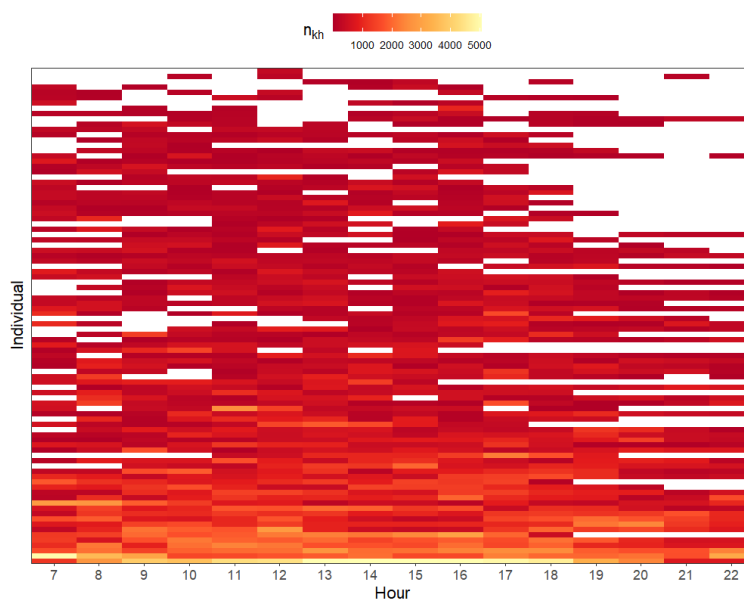
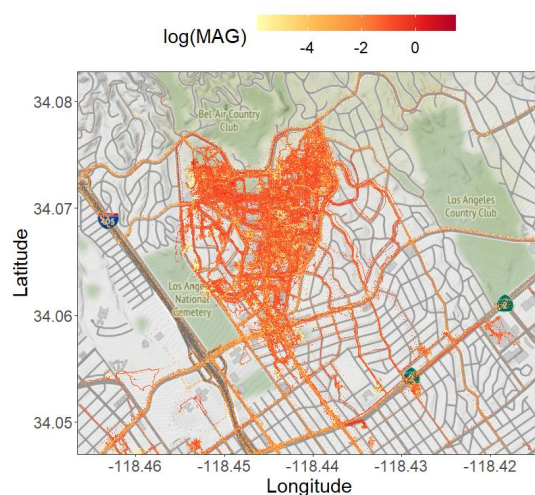Fig 2: Missing data pattern by hour of the day and individual.

Fig 3: Observed locations over the Westwood area.

3. The model. The outcomes corresponding to the $K = 93$ subjects are referenced with respect to the time at which they are recorded and the position in the trajectory. While it is tempting to work with a spatiotemporal process, dependence introduced by such processes may not be appropriate. An individual can visit the same location numerous times in his/her trajectory. These revisits need not occur at regular intervals and can be at distant time points. This suggests that proximity of two spatial locations in a trajectory need not result in strongly dependent MAGs recorded there. It appears more reasonable to model dependence among MAG measurements through a temporal process. In fact, such temporal processes can be motivated by the position vectors defining the trajectories as we describe below.
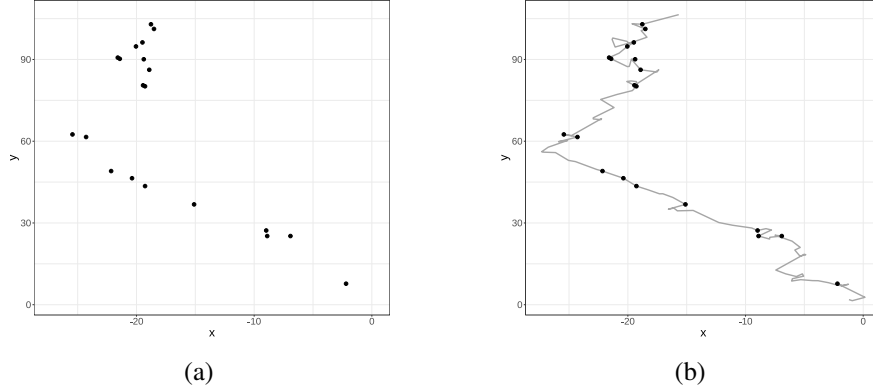
Fig 4: Example of observed points (a) and trajectory (b): black dots are realizations, grey line is domain of the process

Let $Z_k(\cdot) : \mathbb{R}^2 \to \mathbb{R}$ be a spatial process corresponding to individual k. The domain of $Z_k(\cdot)$ is restricted to the trajectories $\gamma_k(t) = (\gamma_k^x(t), \gamma_k^y(t))$, where $k = 1, \ldots, K$ and $t \in \mathbb{R}^+$, which defines the movements of the k-th individual along time. As shown in Figure 4, the process actually belongs to a one-dimensional space, for which we define a proper distance measure $d(t_{ki}, t_{kj}) = \|\gamma_k(t_{kj}) - \gamma_k(t_{ki})\|$, where t_{ki} is the i-th recorded time point from individual k. We approximate such distances as the elapsed time between the two points $d(t_{ki}, t_{kj}) = |t_{kj} - t_{ki}|$, which would result in a good approximation of the spatial distance (especially if the subject is moving at constant speed). More generally, the elapsed separation across time will reflect dependence better than the spatial distance. The faster an individual is moving from one point to the other, the shorter the time elapsed, and higher the correlation between the two measurements. Hence, we model our measurements as $Y_k(\cdot) \equiv Z_k \circ \gamma_k(\cdot) : \mathbb{R}^+ \to \mathbb{R}$, which, by construction, is a valid stochastic process.

This will form the edifice of the model in Section 3.1, where we are modeling the dependence by solely considering stochastic evolution through time. How should spatial information be introduced in the model? Two individuals at the same spatial coordinate experience the same spatial effect but different temporal effects because their physical activities are a function of their trajectory's temporal evolution. An added complication is that trajectories intersect and overlap and, in practice, can have multiple observations at the same location. Even more flexible spatiotemporal covariance kernels (e.g., nonseparable or nonstationarity) will struggle to recognize the above features. Hence, we introduce the spatial effect in the mean using *spline regression* (see Section 3.4).

3.1. *Temporal model.* Let $\mathcal{T} = \cup_{k=1}^K \mathcal{T}_k$ where $\mathcal{T}_k = \{t_{ki}\}_{i=1}^{T_k}$ and $t_{ki} \in \mathbb{R}^+$ be the set of the $n = \sum_{k=1}^K T_K$ observed time points. We model $\boldsymbol{Y}(\mathcal{T})$ as the finite realization of a K-variate process $\boldsymbol{Y}(\cdot)$ over $\mathbb{R}^+$:

$$(3) \qquad \boldsymbol{Y}(t) = \boldsymbol{X}(t, \boldsymbol{\gamma}(t))^\top \boldsymbol{\beta} + \boldsymbol{w}(t) + \boldsymbol{\varepsilon}(t), \quad t \in \mathbb{R}^+,$$

where $\boldsymbol{Y}(t) = (Y_1(t), Y_2(t), \ldots, Y_K(t))^\top$ is a $K \times 1$ vector of measurements at time t on the K individuals, $\boldsymbol{X}(t, \boldsymbol{\gamma}(t))$ is a $p \times K$ matrix, each row being the values of a covariate for the K individuals, $\boldsymbol{w}(t) = (w_1(t), w_2(t), \ldots, w_K(t))^\top$ is a $K \times 1$ vector comprising a temporal process for each individual, and $\boldsymbol{\varepsilon}(t) \sim \mathcal{N}_K(0, \tau^2 \boldsymbol{I}_K)$, $\tau^2 \in \mathbb{R}^+$, is a white noise process for measurement error. Each element of $\boldsymbol{w}(t)$ is specified as $w_k(t) \stackrel{ind}{\sim} \mathcal{GP}(0, c_{\boldsymbol{\theta}}(\cdot, \cdot))$, where $c_{\boldsymbol{\theta}}(\cdot, \cdot)$ is a covariance function with parameters $\boldsymbol{\theta} \in \Theta$.

Let y_{ki} and $\boldsymbol{x}_{ki}$ be the outcome and covariates for individual k at time point t_{ki}, respectively, so $\{(y_{ki}, \boldsymbol{x}_{ki}) : k = 1, \ldots, K, i = 1, \ldots T_k\}$ is the observed data. Let $\boldsymbol{y}_k$ and $\boldsymbol{w}_k$ be $T_k \times 1$ vectors comprising all measurements and random effects on patient k, respectively. Forming the $n \times 1$ vectors $\boldsymbol{y} = \left[\boldsymbol{y}_{1\cdot}^\top \boldsymbol{y}_{2\cdot}^\top \cdots \boldsymbol{y}_{K\cdot}^\top\right]^\top$ and $\boldsymbol{w} = \left[\boldsymbol{w}_{1\cdot}^\top \boldsymbol{w}_{2\cdot}^\top \cdots \boldsymbol{w}_{K\cdot}^\top\right]^\top$, and the $n \times p$ matrix $\boldsymbol{X} = \left[\boldsymbol{X}_{1\cdot}^\top \boldsymbol{X}_{2\cdot}^\top \ldots \boldsymbol{X}_{K\cdot}^\top\right]^\top$, where $\boldsymbol{X}_k$ is the $T_k \times p$ matrix of predictors corresponding to $\boldsymbol{y}_k$, we extend (3) to a hierarchical model with posterior distribution

(4)
$$p(\boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{\theta}, \tau^2 \,|\, \boldsymbol{y}) \propto p(\boldsymbol{\theta}, \tau^2) \times N(\boldsymbol{\beta} \,|\, \boldsymbol{\mu}_\beta, \boldsymbol{V}_\beta) \times N(\boldsymbol{w} \,|\, \boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{\theta}}) \times N(\boldsymbol{y} \,|\, \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{w}, \tau^2 \boldsymbol{I}_n)\,.$$

The covariance matrix $\boldsymbol{C}_{\boldsymbol{\theta}} = \text{diag}\left(\boldsymbol{C}_{\boldsymbol{\theta}_1,1}, \boldsymbol{C}_{\boldsymbol{\theta}_2,2}, \ldots, \boldsymbol{C}_{\boldsymbol{\theta}_K,K}\right)$ is $n \times n$ block-diagonal with $\boldsymbol{C}_{\boldsymbol{\theta}_k,k} = [c_{\boldsymbol{\theta}}(t_{ki}, t_{kj})]$ as the $T_k \times T_k$ temporal covariance matrix corresponding to individual k. Each individual is allowed its own covariance parameters, $\boldsymbol{\theta}_k$, and $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_K\}$ in (4) is the collection of all the covariance kernel parameters. Applying (4) involves the determinant and inverse of $\boldsymbol{C}_{\boldsymbol{\theta}}$, which require $\mathcal{O}(n^2)$ storage space and $\mathcal{O}(n^3)$ floating point operations (flops). The block-diagonal structure of $\boldsymbol{C}_{\boldsymbol{\theta}}$ considerably alleviates this burden since $\det(\boldsymbol{C}) = \prod_{k=1}^K \det(\boldsymbol{C}_{\boldsymbol{\theta}_k,k})$ and $\boldsymbol{C}^{-1} = \text{diag}\left(\boldsymbol{C}_{\boldsymbol{\theta}_1,1}^{-1}, \boldsymbol{C}_{\boldsymbol{\theta}_2,2}^{-1}, \ldots, \boldsymbol{C}_{\boldsymbol{\theta}_K,K}^{-1}\right)$. This reduces the flop count from $\mathcal{O}(n^3) = \mathcal{O}((\sum_{k=1}^K T_k)^3)$ to $\mathcal{O}(K\sum_{k=1}^K (T_k)^3)$, with a significant saving of calculations especially when the T_k's are reasonably small $(< 10^4)$. Furthermore, each $\boldsymbol{C}_{\boldsymbol{\theta}_k,k}$ can be computed in parallel rendering further further scalability to the algorithm.

However, analyzing the Actigraph data in Section 2 will involve $T_k > 10^5$ measurements from some individuals. Full inference will be impractical without any exploitable structure for each $\boldsymbol{C}_{\boldsymbol{\theta}_k,k}$. Analyzing massive spatiotemporal data has witnessed burgeoning interest and a comprehensive review is beyond the scope of this work (see, e.g., Banerjee, 2017; Heaton et al., 2019, and references therein). We will pursue an approximation due to Vecchia (Vecchia, 1988) that has generated substantial recent interest (Datta et al., 2016a,b; Katzfuss et al., 2020; Katzfuss and Guinness, 2021; Peruzzi, Banerjee and Finley, 2022) in scalable Bayesian modeling.

3.2. *Independent DAG models over individuals.* We adapt *Vecchia's* likelihood approximation (Vecchia, 1988) to the random effects $\boldsymbol{w}_k$ for each $k = 1, 2, \ldots, K$. Beginning with the observed time points $\{t_{k1} < t_{k2} < \cdots < t_{kT_k}\}$ for individual k and the directed acyclic graphical (DAG) representation $p(\boldsymbol{w}_k) = p(w_{k1}) \prod_{i=2}^{T_k} p(w_{ki} \,|\, w_{k1}, \ldots, w_{k(i-1)})$, we define

(5)
$$p(\boldsymbol{w}_k) \approx \widetilde{p}(\boldsymbol{w}_k) = p(w_{k1}) \prod_{i=2}^{T_k} p(w_{ki} | \boldsymbol{w}_{k,N(i)})\,,$$

where $\widetilde{p}(\cdot)$ is the joint density derived from $p(\boldsymbol{w}_k)$ by restricting the parents (conditional sets) of each w_{ki} in the DAG to a set $w_{kN(i)} = \{w_{kj} : j \in N(i)\}$, where $N(i)$ is a set of prefixed size m comprising the m nearest neighbors of t_{ki} from the past. Thus, $N(i) = \{t_{k(i-m)}, \ldots, < t_{k(i-1)}\}$ for $i > m$ and $N(i) = \{t_{k1}, \ldots, t_{k(i-1)}\}$ for $i \leq m$. Such approximations yield valid probability likelihoods (Lauritzen, 1996; Stein, Chi and Welty, 2004; Murphy, 2012) and can be extended to stochastic processes (Datta et al., 2016a) for inference on arbitrary time points.

The connection between sparsity and conditional independence follows by writing (5) as a linear model $\boldsymbol{w}_k = \boldsymbol{A}_k \boldsymbol{w}_k + \boldsymbol{\eta}_k$, where $\boldsymbol{A}_k$ is a $T_k \times T_k$ strictly lower triangular matrix, $\boldsymbol{\eta}_k \sim \mathcal{N}_{T_k}(\boldsymbol{0}, \boldsymbol{D}_k)$ and $\boldsymbol{D}_k$ is the $T_k \times T_k$ diagonal matrix such that $[\boldsymbol{D}_k]_{ii} = d_{ii} = \text{Var}\,(w_{ki}|\{w_{kj}, j < i\})$ for $i = 1, \ldots, T_k$. The DAG imposes the lower-triangular structure on $\boldsymbol{A}_k$ and its (i,j)-th entry is allowed to be nonzero only for $j \in N(i)$. Therefore, each row of $\boldsymbol{A}_k$ has at most m nonzero entries so that $\widetilde{\boldsymbol{C}}_k^{-1} = (\boldsymbol{I}_{T_k} - \boldsymbol{A}_k)^\top \boldsymbol{D}_k^{-1} (\boldsymbol{I}_{T_k} - \boldsymbol{A}_k)$ is sparse,

where $\widetilde{\boldsymbol{C}}_k^{-1}$ is the precision matrix corresponding to $\widetilde{p}(\boldsymbol{w}_k)$. Replacing $\boldsymbol{C}$ with $\widetilde{\boldsymbol{C}}$ in (4) yields a computationally efficient hierarchical model with $\prod_{k=1}^K N(\boldsymbol{w}_k \,|\, \boldsymbol{0}, \widetilde{\boldsymbol{C}}_k)$ as the prior on $\boldsymbol{w}$.

The key observation is that the nonzero elements of the i-th row of $\boldsymbol{A}_k$ is the solution $\boldsymbol{a}_k$ of the $m \times m$ linear system $\boldsymbol{C}_{\boldsymbol{\theta},k}[N(i), N(i)]\boldsymbol{a}_k = \boldsymbol{C}_{\boldsymbol{\theta},k}[N(i), i]$, where $[\cdot, \cdot]$ indicates submatrices defined by the given row and column index sets. Obtaining the nonzero elements of $\boldsymbol{A}_k$ and $\boldsymbol{D}_k$ costs $\mathcal{O}(T_k m^3)$ (scales linearly with T_k) instead of $\mathcal{O}(T_k^3)$ as would have been without sparsity. This cheaply delivers the quadratic form $\boldsymbol{w}_k^\top \widetilde{\boldsymbol{C}}_k^{-1} \boldsymbol{w}_k$ in terms of $\boldsymbol{A}_k$ and $\boldsymbol{D}_k$ and the determinant $\det(\widetilde{\boldsymbol{C}}_k) = \prod_{i=1}^{T_k} d_{ii}$ at almost no additional cost. The lower triangular matrix $\boldsymbol{A}_k$ is not just sparse but also banded, with a lower bandwidth equal to m. Consequently, $\widetilde{\boldsymbol{C}}_k^{-1}$ is also banded with lower and upper bandwidth equal to m. This leads to further accrual of computational benefits. The overall cost is $\mathcal{O}(\sum_{k=1}^K T_k m^3) = \mathcal{O}(nm^3)$ (linear in n) for computing the posterior for any given values of the parameters.

3.3. *Implementation using collapsed models.* The Bayesian hierarchical model in (4), either with $\boldsymbol{C}_{\boldsymbol{\theta}}$ or with $\widetilde{\boldsymbol{C}}_{\boldsymbol{\theta}}$ in the prior for $\boldsymbol{w}$, allows full posterior inference for $\{\boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{\theta}, \tau^2\}$ using Markov chain Monte Carlo (MCMC). Gibbs sampling with random walk Metropolis steps provide full conditional distributions in closed form for $\{\boldsymbol{\beta}, \boldsymbol{w}\}$ and also for τ^2 with an $\mathcal{IG}(a_\tau, b_\tau)$ prior. However, this convenience is nullified in practice by strong autocorrelation and poor mixing of the chains (Liu, Wong and Kong, 1994). Samplers based on spatial DAG-based models have been devised, explored and compared in Finley et al. (2019). Instead of (4), we sample from

$$(6) \qquad p(\boldsymbol{\beta}, \boldsymbol{\theta}, \tau^2 \,|\, \boldsymbol{y}) \propto p(\boldsymbol{\theta}, \tau^2) \times N(\boldsymbol{\beta} \,|\, \boldsymbol{\mu}_\beta, \boldsymbol{V}_\beta) \times N(\boldsymbol{y} \,|\, \boldsymbol{X}\boldsymbol{\beta}, \widetilde{\boldsymbol{C}}_{\boldsymbol{\theta}} + \tau^2 \boldsymbol{I}_n) \,,$$

which is derived from (4) by integrating out $\boldsymbol{w}$, thereby "collapsing" the parameter space to a much smaller domain without $\boldsymbol{w}$. This considerably improves mixing and convergence.

We will need to compute the inverse and determinant of $\widetilde{\boldsymbol{\Lambda}} = \widetilde{\boldsymbol{C}}_{\boldsymbol{\theta}} + \tau^2 \boldsymbol{I}_n$, which is $n \times n$. While $\widetilde{\boldsymbol{\Lambda}}^{-1}$ does not share the same convenient factorization of $\widetilde{\boldsymbol{C}}^{-1}$ and is also not guaranteed to be sparse, the Sherman-Woodbury-Morrison formulas reveal

$$(7) \qquad \widetilde{\boldsymbol{\Lambda}}^{-1} = \tau^{-2}\boldsymbol{I} - \tau^{-4}\boldsymbol{\Omega}^{-1}, \quad \text{with} \quad \boldsymbol{\Omega} = \widetilde{\boldsymbol{C}}^{-1} + \tau^{-2}\boldsymbol{I} \,,$$

where $\boldsymbol{\Omega}$ enjoys the same sparsity as $\boldsymbol{C}^{-1}$. Moreover, $\det(\widetilde{\boldsymbol{\Lambda}}) = \tau^{2n}\det(\widetilde{\boldsymbol{C}})\det(\boldsymbol{\Omega})$. The core of the algorithm is therefore to compute $\widetilde{\boldsymbol{\Lambda}}^{-1}$ through $\boldsymbol{\Omega}$. In our application, the random effect is assumed to be the realization of K independent temporal processes. As discussed in Section 3.2, this implies a block-diagonal structure for $\widetilde{\boldsymbol{C}}$ that can be shown to be shared also by $\boldsymbol{\Omega}$ (see Eq. (7)). Each block $\boldsymbol{\Omega}_k$ of $\boldsymbol{\Omega}$ can be computed independently for each individual and the same holds for its inverse and its determinant. This means that the body of the algorithm will consist of a loop over all the individuals, which allows for straightforward parallelization (see Algorithm 1 in the Supplementary Material (Alaimo Di Loro et al., 2023)). Unlike in spatial DAGs (Datta et al., 2016a; Finley et al., 2019), we do not need fill-reducing permutation methods since neighbors sets for temporal processes consist of contiguous observations and $\{\boldsymbol{\Omega}_k\}_{k=1}^K$ are banded matrices with no gaps.

We devised a Gibbs sampler with Metropolis random walk updates for (6), where $\boldsymbol{\beta}$ is updated from its full conditional distribution, while $\{\boldsymbol{\theta}, \tau^2\}$ are updated using an adaptive Metropolis step based on Haario et al. (2001). Here, after the first few iterations, a new proposal covariance matrix is regularly computed on the run according to the empirical covariance of the current chain. Subsequently, a mixture of the original and adaptive proposal is used as the new proposal. Convergence toward the desired acceptance rate is assured for

an appropriate choice of the variance terms and of the adaptation rule (Roberts and Rosenthal, 2009). The algorithm has been coded using the `R 4.0.5` statistical environment. All expensive computations are managed by the `Eigen` library (version 3.3.7), which provides efficient routines for numerical linear algebra with an emphasis on sparse matrices. Our implementation of (6) outperforms the algorithms that update $\boldsymbol{w}$ in terms of computational speed as it is implemented in the `spNNGP` package (Finley, Datta and Banerjee, 2017). We present these comparisons in the Supplementary Materials (Alaimo Di Loro et al., 2023) including a link to the GitHub repository hosting codes to implement the models.

3.4. *Including spatial effects.* Accounting for spatial information in our Actigraph dataset presents some new considerations. As mentioned in Section 1, spatial information is available to us in terms of the physical location along the trajectory as well as through covariates that are functions of space. Considering the discussion in Section 3, the analytical goals of this dataset suggest accounting for spatial heterogeneity. Here, as argued earlier, modeling $\boldsymbol{w}(\cdot)$ in (3) as a spatio-temporal process, including scalable versions, has challenges given that: (i) the trajectory's domain does not have a positive area; and (ii) associations among the measurements are more amenable to the temporal scale. Therefore, we introduce spatial effects into the mean employing a smooth function of space, $f_S(\cdot) : \mathbb{R}^2 \to \mathbb{R}$, approximated by a spline basis representation (see, e.g., Goodman and Hardin, 2006; Ramsay and Silverman, 2007). For instance, if J_x and J_y are the dimensions of independently defined B-spline basis expansions on the x and y coordinates, respectively, then $f_S((x,y)) \approx \widetilde{f}_S((x,y)) = \sum_{j_X=1}^{J_X} \sum_{j_Y=1}^{J_Y} \beta_{S,(j_X,j_Y)} B_{x,j_x}(x) B_{y,j_y}(y)$, where $B_{x,j_X} = [\boldsymbol{B}_x]_{j_x}$ and $B_{y,j_Y} = [\boldsymbol{B}_y]_{j_y}$ are the j_x-th and j_y-th element of the B-spline basis along the two axis. For any location $(x,y) \in \mathbb{R}^2$ the elements of the previous sum can be more compactly expressed through the tensor product basis $\boldsymbol{B}_S(x,y) = (\boldsymbol{B}_x \otimes \boldsymbol{B}_y)(x,y)$. The size of this basis is $J_S = J_x \cdot J_y$ and depends on the size of the two original spline basis, which in turn depends on the chosen number of knots $knots_x, knots_y$ and degree deg_x, deg_y (namely $J_c = knots_c + deg_c$ for $c = x, y$). We now modify (3) to include the spline,

$$(8) \qquad \boldsymbol{Y}(t) = \boldsymbol{X}(t, \boldsymbol{\gamma}(t))\boldsymbol{\beta} + \boldsymbol{B}_S(\boldsymbol{\gamma}(t))\boldsymbol{\beta}_S + \boldsymbol{w}(t) + \boldsymbol{\varepsilon}(t), \qquad t \in \mathbb{R}^+,$$

where $\boldsymbol{\gamma}(t) = \{\boldsymbol{\gamma}_1(t), \boldsymbol{\gamma}_2(t), \ldots, \boldsymbol{\gamma}_K(t)\}$, $\boldsymbol{\gamma}_k(t) = (\gamma_{k,x}(t), \gamma_{k,y}(t)) : \mathbb{R}^+ \to \mathbb{R}^2$ is the trajectory function mapping time t for individual k to its position and $\boldsymbol{B}_S(\boldsymbol{\gamma}(t))$ is the $K \times J_S$ matrix with row k corresponding to the J_S basis elements for the coordinates at time point t for individual k. A proper choice of J_S (i.e. knots and degree) is required to fit a spline surface flexible enough to describe the spatial variations at the scale of interest without incurring over-fitting. Let $\boldsymbol{B} = \boldsymbol{B}_S(\boldsymbol{\gamma}(\mathcal{T}))$ be the $n \times J_S$ matrix containing the B-spline basis elements evaluated at the observed location of each individual $\boldsymbol{\gamma}(\mathcal{T}) = \{\gamma_1(t_{11}), \gamma_1(t_{12}), \ldots, \gamma_K(t_{KT_k})\}$. Following Equation (6), we sample from
$$(9)$$
$$p(\boldsymbol{\beta}, \boldsymbol{\beta}_S, \boldsymbol{\theta}, \tau^2 \,|\, \boldsymbol{y}) \propto p(\boldsymbol{\theta}, \tau^2) \times p_S(\boldsymbol{\beta}_S) \times N(\boldsymbol{\beta} \,|\, \boldsymbol{\mu}_\beta, \boldsymbol{V}_\beta) \times N(\boldsymbol{y} \,|\, \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{B}\boldsymbol{\beta}_S, \widetilde{\boldsymbol{C}}_{\boldsymbol{\theta}} + \tau^2 \boldsymbol{I}_n),$$

where the prior $p_S(\cdot)$ needs to be specified. The Actigraph data includes millions of observations in a limited study area, of which some assume different values in the same location (or in its immediate vicinity) so over-fitting will not be an issue. However, some areas present sparsely observed points (trajectories are not uniformly distributed, as shown in Figure 3). This may cause coefficients corresponding to those regions to be weakly identified. To control for the balance of all these components, we may assign ad-hoc priors to the spatial spline regression coefficients (Eilers and Marx, 1996) for penalizing deviation from a certain degree of smoothness and favoring identifiability. This behavior suggests the Bayesian P-Spline (Hastie et al., 2000; Lang and Brezger, 2004). While keeping the Gaussian priors, we

effectuate shrinkage by choosing a suitable precision matrix $\boldsymbol{P}$ and introducing a shrinkage parameter λ at a deeper level of the hierarchy. To be precise, $\boldsymbol{\beta}_S \,|\, \lambda \propto \exp\left\{ -\frac{\lambda}{2} \cdot \boldsymbol{\beta}_S \boldsymbol{P} \boldsymbol{\beta}_S^\top \right\}$ and $\lambda \sim \mathcal{G}(\alpha_\lambda, \beta_\lambda)$. We consider two possible forms for $\boldsymbol{P}$, which imply different penalization for the coefficients:

- **Ridge-like** prior, which is to say $\boldsymbol{P} = \boldsymbol{P}_{RL} = \boldsymbol{I}_{J_S}$;
- **First-order random walk** prior, which is to say:

$$\boldsymbol{P} = \boldsymbol{P}_{RW} : [\boldsymbol{P}_{RW}]_{ij} = \begin{cases} n_i & i = j \\ -1 & i \sim j \\ 0 & \text{otherwise} \end{cases}$$

where n_i is the number of neighbors of knot i and $i \sim j$ denotes a neighboring relationship between the knots.

Both precision matrices provide a multivariate Gaussian prior distribution on the coefficients. However, the latter is improper since $\text{rank}\,(\boldsymbol{P}_{RW}) < J_S$. Nevertheless, if we collect the B-Spline basis elements with the other covariates as $\boldsymbol{X}^* = [\boldsymbol{X}, \boldsymbol{B}]$ and stack the corresponding coefficients into the joint vector $\boldsymbol{\psi} = [\boldsymbol{\beta}, \boldsymbol{\beta}_S]$, then the posterior distribution of the latter is a proper multivariate Gaussian with full conditional distribution $\boldsymbol{\psi} \,|\, \cdot \propto \mathcal{N}_J\left(\boldsymbol{\psi} \,\big|\, \boldsymbol{G}^{-1} \boldsymbol{g}, \, \boldsymbol{G}^{-1} \right)$, where $\boldsymbol{G} = \boldsymbol{X}^{*\top} \widetilde{\boldsymbol{\Lambda}}^{-1} \boldsymbol{X}^* + \boldsymbol{V}_{\boldsymbol{\psi}}^{-1}$ and $\boldsymbol{g} = \boldsymbol{X}^{*\top} \widetilde{\boldsymbol{\Lambda}}^{-1} \boldsymbol{y} + \boldsymbol{V}_{\boldsymbol{\psi}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\psi}}$ with $\boldsymbol{V}_{\boldsymbol{\psi}}^{-1} = \text{diag}\left(\boldsymbol{V}_{\boldsymbol{\beta}}^{-1}, \lambda \cdot \boldsymbol{P} \right)$ and $\boldsymbol{g} = \left[\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\mu}_{\boldsymbol{\beta}_S} \right]^\top = \boldsymbol{0}^\top$. Moreover, the Gamma prior on λ implies a Gamma full-conditional distribution $\lambda \,|\, \cdot \propto \mathcal{G}\left(\lambda \,\big|\, \alpha_\lambda + 1/2, \, \beta_\lambda + \boldsymbol{\beta}_S^\top \boldsymbol{P} \boldsymbol{\beta}_S \right)$.

Estimating the model in (8) is achieved through a straightforward extension of Algorithm 1. We jointly update $\boldsymbol{\psi}$ and λ from their full conditional distributions. In particular, the Gibbs' sampling step of Algorithm 1 can be modified to get full inference also on the spline coefficients $\boldsymbol{\beta}_S$ and the shrinkage parameter λ (see Algorithm 2 in the Supplementary Material Alaimo Di Loro et al., 2023). In practical terms, this requires J_S additional linear coefficients to be estimated, whose size $p^* = p + J_S$ may undermine the efficiency of the algorithm. For example, calculations in Step 1b are quadratic w.r.t. $p^* \to \mathcal{O}(np^{*2})$. Steps 1a and 1b (i.e. the most expensive in p^*) are executed in the first iteration and subsequently, only in those iterations where new values of $\boldsymbol{\theta}$ are accepted. When $\boldsymbol{\theta}$ is rejected, we retain in memory the previously computed value (which would stay unchanged). Thus, if we attain an optimal acceptance rate of $\approx 20\% - 30\%$ in the Metropolis Hastings step on $\boldsymbol{\theta}$, the computation is avoided in the majority of cases with a sensible improvement in computation time and speed.

3.5. *Simulations.* We conducted simulation experiments to evaluate the model described in Section 3.4 and compared the performance of our algorithm in terms of fitting, prediction error and computational speed with other routines available from the spNNGP package (Finley, Datta and Banerjee, 2017). Additional comparative experiments are provided in the Supplementary Material (Alaimo Di Loro et al., 2023). We executed our MCMC algorithms on a computing environment equipped with 12 modern computational nodes with 32 cores each, roughly equivalent to 3 TeraFlop/sec, and 256Gb of RAM. Each of the presented applications have been executed on a single node exploiting the computational power of all cores. The results presented are based upon posterior samples that were retained after diagnosing convergence using visual tools (e.g., traceplots, autocorrelation), effective sample sizes, Monte Carlo standard errors (MCSE) and other diagnostics offered by the coda, mcse and bayesplot packages in the R computing environment; the Supplementary Material (Alaimo Di Loro et al., 2023) includes specific details.

Param. (True)	S-Spline		P-Spline	
	Point	Interval	Point	Interval
β_{01} (-3.76)	-3.799	(-3.846,-3.752)	-3.797	(-3.844,-3.75)
β_{02} (0.65)	0.572	(0.523,0.62)	0.575	(0.526,0.623)
β_{03} (-0.60)	-0.649	(-0.697,-0.6)	-0.646	(-0.693,-0.598)
β_{04} (2.36)	2.326	(2.277,2.374)	2.328	(2.28,2.376)
β_{05} (-0.33)	-0.359	(-0.408,-0.31)	-0.356	(-0.404,-0.308)
β_{1} (2.59)	2.599	(2.59,2.608)	2.599	(2.59,2.608)
β_{2} (2.70)	2.691	(2.683,2.7)	2.691	(2.683,2.7)
β_{3} (-0.58)	-0.586	(-0.595,-0.577)	-0.586	(-0.595,-0.577)
σ^2 (1)	1.001	(0.973,1.032)	0.993	(0.965,1.023)
ϕ (1)	0.994	(0.948,1.04)	1.01	(0.964,1.063)
τ^2 (1)	1.001	(0.984,1.018)	1.001	(0.984,1.018)
Metric	Out-of-sample	In-sample	Out-of-sample	In-sample
Coverage	0.95	0.99	0.95	0.99
RMSPE (r)	0.07 (1.18)	0.03 (0.84)	0.07 (1.19)	0.03 (0.84)
PIW	4.66	4.44	4.66	4.44
DIC	115'543		115'556	
Fitting time (h)	2.18		2.2	

TABLE 1

Parameter estimates, predictive validation and fitting times (hours) on the simulated dataset for all the considered models.

We first generated $T_k = 2 \times 10^5$ time points for $K = 5$ individuals, where each time point t_{ki} followed exponential waiting times between observations, i.e. $t_{ki} = \sum_{h=1}^{i-1} \delta_h$, and $\delta_h \overset{iid}{\sim} \text{Exp}(5)$. Given the time points, we constructed spatial trajectories $\gamma_k(\cdot)$, $k = 1, \ldots, K$, by simulating $\boldsymbol{s}_k = [\gamma_k(t_{k1}), \ldots, \gamma_k(t_{kT_k})]^\top$, where subsequent components were independent Gaussian random walks over the square $\mathcal{S} = (1, 10) \times (1, 10)$, with the variance of each step along the horizontal and vertical axis proportional to the elapsed time between two subsequent observations. If the trajectory left the square, it was projected onto the border and the next step would resume from there. The simulated trajectories are shown in Figure 5a.

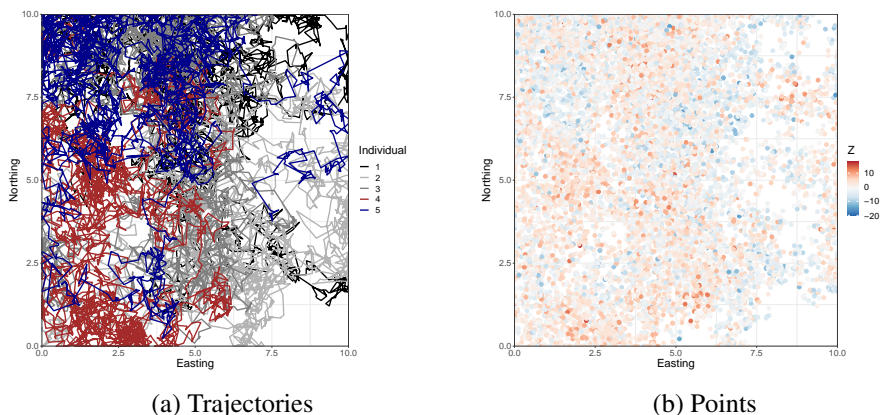(a) Trajectories

(b) Points

Fig 5: Observed trajectories (a) and observed points (b) for the simulated dataset.

Given the time points and positions (Figure 5b), we generated the latent temporal Gaussian processes $w_k(\cdot) \overset{ind}{\sim} \mathcal{GP}(0, c_{\boldsymbol{\theta}}(\cdot, \cdot))$ with an exponential covariance $c_{\boldsymbol{\theta}}(t, t') = \sigma^2 \exp\{-\phi \cdot |t - t'|\}$, where $\sigma^2 > 0$ represents the variance of the process, $\phi > 0$ is the decay in temporal correlation (range) and $\tau^2 > 0$ the residual variance (nugget). The spatial effects are then introduced through $f_S(\cdot) : \mathcal{S} \to \mathbb{R}$ by considering a tensor product spline basis of degree 2 and with 9 knots over the square domain (including boundary knots), where the spline coef-

ficients $\boldsymbol{\beta}_S$ have been fixed to randomly generated values from $\mathcal{N}_{81}\left(\mathbf{0}, \lambda \boldsymbol{I}_{81}\right)$ with $\lambda = 0.5$. The model also included individual-specific intercepts $\{\beta_{0k}\}_{k=1}^5$ and the effect of 3 covariates with random values drawn independently at each location from a $\mathcal{N}(0,1)$ distribution, leading to covariate vectors $\{\boldsymbol{x}_{ki}\}_{i=1}^{T_k}$, $k = 1, \dots, K$. The effect of the covariates is assumed common across individuals, and set to be determined by slopes $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3]^\top$.

We generated values of the outcome for individual k at time t_{ki} and location $\boldsymbol{s}_{ki} = \gamma_k(t_{ki})$ according to the generative process defined by (8) with parameters fixed as above. This yielded a simulated dataset $D_{sim} = \left\{(\text{Ind}_j, t_j, \boldsymbol{s}_j, y_j, \boldsymbol{x}_j^\top)\right\}_{j=1}^n$ with $n = 10^5$ observations, where Ind_j denotes the individual corresponding to row j. Then, we fit the model in (9) on 70% of the total observations in D_{sim} using Algorithm 1 with the Gibbs' sampling modified as in Algorithm 2. The remaining 30% were held out to assess out-of-sample predictive performances in terms of *Relative and Root Mean Squared Prediction Error* (RMSPE), *Coverage*, and *Predictive Interval Width* (PIW). Intercept and slope regression parameters were assigned $\mathcal{N}(0, 10^6)$ priors; the variance components, σ^2 and τ^2, were both assigned inverse Gamma $\mathcal{IG}(2,2)$ priors; and the decay parameter ϕ received a Gamma prior $\mathcal{G}(1,1)$. For the spline coefficients, we considered both the penalized versions in Section 3.4. The first is referred to as an S-Spline (shrinking splines), and the second as P-Spline (penalized splines).

Table 1 presents the posterior estimates. We also included the Deviance Information Criterion (DIC) for both models. Performances in the two settings are almost identical, but the DIC favors the S-Spline model. This is not surprising as the data were generated using an analogous shrinkage prior for the $\boldsymbol{\beta}_S$'s. Further details, including the estimates of the spline coefficients are provided in the Supplementary Material (Alaimo Di Loro et al., 2023). Figure 6 presents the posterior estimate of the spatial surface. We compare the true latent surface with just the S-Splines as it performs slightly better with respect to the DIC, but notice that P-Splines provides practically identical estimates.
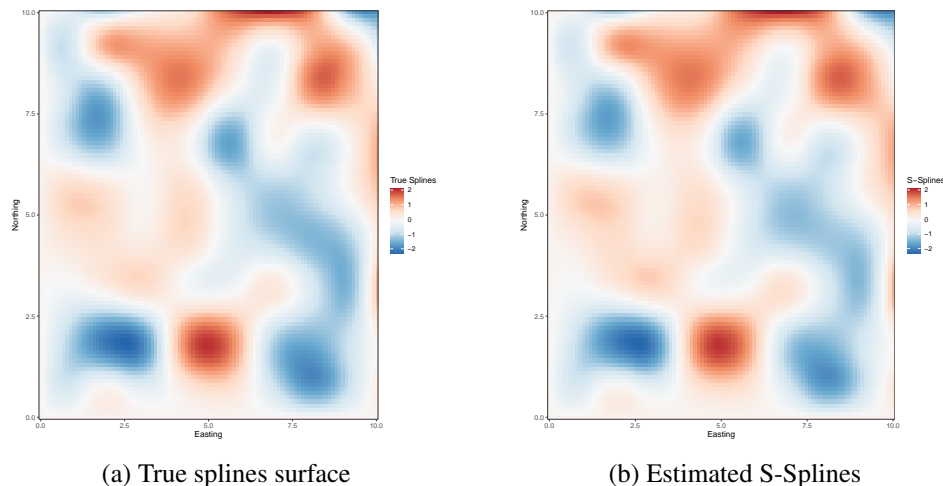

(a) True splines surface   (b) Estimated S-Splines

Fig 6: True and estimated spline surfaces using S-Splines.

4. Application. We apply the proposed model in (8) to estimate the MAG (measured in G) for participants in the study accounting for subject-specific features and spatial effects on the mean, while modeling the latent temporal dependence as described in Section 3.1. We split the data into training (70%) and testing (30%) subsets, where the records have been allocated to each subset according to a random sample stratified by individual. The testing

set is used to assess the out-of-sample predictive performances in terms of *Relative Mean Squared Prediction Error* (RMSPE), *root Mean Squared Prediction Error* (rMSPE), *Coverage* (Cov), and *Predictive Interval Width* (PIW). Posterior inferences are based on $5,000$ samples retained after diagnosed convergence from $10,000$ MCMC iterations.

4.1. *Model specification.* Spatial effects are introduced by considering the tensor product of two analogous univariate B-spline basis on each spatial axis. After a preliminary validation through the DIC (see Figure 14 of the appendix), we choose two bases of degree 3 with 12 equally spaced knots over a square encompassing Westwood. This sums up to $J_S = (9+3) \times (9+3) = 144$ terms for our complete spline basis, including the boundary knots. All numerical variables in $X(\cdot, \cdot)$ have been standardized for improving the efficiency of the MCMC sampling (Gilks and Roberts, 1996). The presence of temporal dependence in individual trajectories was investigated through an individual-specific exploratory analysis on the residuals from a standard linear regression and an Ornstein-Uhlenbeck process (GP with an exponential covariance function) was specified to capture temporal dependence as a parsimonious and effective model for the behavior of the underlying residual process.

Finally, the outcome is log-transformed in order to comply with the Gaussianity assumption of the model. We denote the parameter associated with variable "varname" as β_{varname} and the levels of each categorical covariate as $\text{varname}_{(j)}$ for $j = 1, \ldots, J_{\text{varname}}$. Hence,

(10)

$$
\mathbb{E}\left[\log(MAG_k(t))\right] = \beta_0 + \sum_{j=2}^{J_{\text{BMI}}} \beta_{\text{BMI},j} \cdot \mathbb{I}\left(\text{BMI}_k = \text{BMI}_{(j)}\right) + \sum_{j=2}^{J_{\text{Sex}}} \beta_{\text{Sex},j} \cdot \mathbb{I}\left(\text{Sex}_k = \text{Sex}_{(j)}\right) +
$$

$$
+ \sum_{j=2}^{J_{\text{Age}}} \beta_{\text{Age},j} \cdot \mathbb{I}\left(\text{Age}_k = \text{Age}_{(j)}\right) + \sum_{j=2}^{J_{\text{Eth}}} \beta_{\text{Eth},j} \cdot \mathbb{I}\left(\text{Eth}_k = \text{Eth}_{(j)}\right) +
$$

$$
+ \beta_{\text{distHome}} \cdot \text{distHome}_k(\gamma_k(t)) + \beta_{\text{NDVI}} \cdot \text{NDVI}(\gamma_k(t)) +
$$

$$
+ \beta_{\text{distParks}} \cdot \text{distParks}(\gamma_k(t)) + \beta_{\text{Slope}} \cdot \text{Slope}(\gamma_k(t)) +
$$

$$
+ \sum_{j=1}^{J_S} \beta_{S,j} B_{S,j}\left(\gamma_k(t)\right)
$$

where $\mathbb{I}(\cdot)$ denotes the indicator function, $w_k(\cdot)$ is the DAG-based approximation (Section 3.2) for $\mathcal{GP}(0, c_{\boldsymbol{\theta}}(\cdot, \cdot))$, and $\epsilon_k(t) \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \tau^2\right)$. The baseline subject represents an underweight Asian female less than 18 years of age. Other socioeconomic factors (e.g. education and income level) have been excluded from the analysis as they are strongly associated with ethnicity and age.

4.2. *Prior distributions.* The prior choices for each set of parameters and/or coefficients followed ad-hoc strategies. We incorporated priors such as $\boldsymbol{\beta} \sim \mathcal{N}_J\left(\mathbf{0}, 10^6 \cdot \boldsymbol{I}_J\right)$, $\sigma^2 \sim \mathcal{IG}(2,2)$ and $\tau^2 \sim \mathcal{IG}(2,2)$ with J being the total number of $\boldsymbol{\beta}$ coefficients. The high spatial density of observations in several areas of the map enables robust estimation of the spatial effects. However, over-fitting may emerge from the the high dimension of the spline basis. Furthermore, there are areas in Westwood that present sparsely observed data-points and the model could struggle to identify the spline coefficients referred to those areas and jeopardize convergence of the MCMC algorithm. Hence, we consider the S-Spline (Ridge-like prior) described in Section 3.4 to mitigate these potential issues, where the shrinkage parameter λ has been assigned a $\mathcal{G}(1,1)$ prior.

4.3. *Results.* Fitting the model required ≈ 7 hours on a computer equipped with 2 processors AMD EPYC 7452, each one having 32 cores for a maximum of 256Gb of RAM. The acceptance rate obtained is $\approx 28\%$, supporting the consistency of our adaptive strategy. Table 2 presents parameter estimates and performance metrics for the model in (8) with the explanatory variables specified in (10) alongside estimates from a Bayesian linear regression model that includes the spatial spline terms, but neglects the temporal dependence structure.

The estimates from the two models are largely consistent with each other, although accounting for temporal dependence tends to somewhat mitigate the effects of some predictors. We anticipate the temporal process to absorb the impact of certain predictors — especially when their relationship with the dependent variable is complex and nonlinear — and this appears to be the case with "Slope", which loses its significant positive impact on MAG once the temporal process is incorporated. All other variables seem to retain the nature of their impact on MAG. These coefficients are interpreted with respect to the baseline measure of an underweight 18 year old Asian female.

The intercept represents the natural logarithm of the MAG for the baseline subject and reckons with both active as well as inactive time points for the high-resolution digitally filtered data. Since we model continuously over time, the resulting value of the MAG, which is about $\exp(-2.9) \approx 0.06$, corresponds to the low region of moderate physical activity level (see the Supplementary Material, Alaimo Di Loro et al., 2023). Other gleanings from Table 2 indicate that MAGs vary by ethnicity in the study cohort as Whites tend to record lower MAGs, while Latin Americans and African Americans tend to register larger MAGs compared to Asians. Impact of Age groups on MAGs also tend to vary with the groups of 25-34 and 45-70 year old subjects tending to register lower MAGs than the baseline (young), while the middle-aged group tends to be higher than the baseline. This is not entirely surprising because subjects in the 25-34 year old group tend to exercise less than the younger (baseline) and middle-aged groups with 25-34 year old subjects having less time as they embark on their careers (less free time), while those in the 45-70 range also tend to follow a less vigorous lifestyle regimen due to their age. The effect of Body Mass Index (BMI) is also seen to vary based upon the categories of weight. While all three categories indicate a significantly higher MAG compared to the baseline, the impact of the overweight, but not obese, category seems higher than the other two. We do not fully know the extent to which larger body weights affect accelerometer readings, but this variation must also account for the fact that higher BMI may also correspond to muscular (not unfit) individuals engaging in more vigorous lifestyle regiments. The spatially-indexed predictors indicated the expected positive impact of NDVI (more greenness encourages more outdoor activities and exercise) while it is also expected, especially in Westwood, that subjects tend to exercise along paths closer to their home thereby explaining the negative coefficient for the weighted distance to home.

The estimate of the temporal decay parameter ϕ implies that the temporal correlation drops to 0.05 in about $3/\hat{\phi} \approx 4.3$ minutes, where $\hat{\phi} \approx 0.7$ is the posterior median of ϕ. Unsurprisingly, including the spatial effect and the temporal process improves predictive performances (RMSPE or PIW in Table 2) over a model including only spatial effects (excluding the temporal process). The model incorporating the temporal process delivers satisfactory coverage and outperforms its competitor in all of the other indices for the training and testing data.

Figure 7a shows the estimated spatial surface, while Figure 7b presents the width of the posterior predictive intervals. The map clearly evinces zones (darker shades of red highlighted with white contours) that tend to depict high levels of physical activity. For example, the largest dark red blob in the north center-left almost perfectly tracks the UCLA campus boundary reflecting a campus environment with active mobility (walking, running, biking). Other zones of high activity identify with locations where more participants in the study live, including those residing in student dorms (northwest corner) and residential areas immedi-

Param.	Model (8) without temporal process Point	Interval	Model (8) with temporal process Point	Interval
Intercept	-2.750	(-2.754, -2.746)	-2.92	(-2.94, -2.91)
Eth. White	-0.128	(-0.146, -0.111)	-0.190	(-0.258, -0.125)
Eth. Other	0.122	(0.110, 0.134)	0.128	(0.077, 0.178)
Eth. Latin-American	0.259	(0.247, 0.271)	0.314	(0.264, 0.362)
Eth. Black/African/Caribbean	0.263	(0.248, 0.278)	0.400	(0.340, 0.461)
Sex Male	-0.348	(-0.358, -0.338)	-0.298	(-0.338, -0.258)
Normal weight	0.121	(0.110, 0.132)	0.297	(0.252, 0.343)
Over weight	0.351	(0.330, 0.372)	0.482	(0.398, 0.566)
Obese	0.220	(0.181, 0.258)	0.401	(0.241, 0.560)
Age (25-34]	-0.387	(-0.398, -0.377)	-0.320	(-0.362, -0.279)
Age (34-45]	0.080	(0.064, 0.097)	0.125	(0.057, 0.191)
Age (45-70]	-0.105	(-0.132, -0.079)	-0.091	(-0.192, 0.006)
Dist. from home	-0.135	(-0.142, -0.128)	-0.074	(-0.102, -0.046)
Slope	0.052	(0.047, 0.0.56)	-0.003	(-0.12, 0.005)
Dist. to parks	-0.221	(-0.227, -0.214)	-0.066	(-0.089, -0.043)
NDVI	0.226	(0.221, 0.231)	0.010	(0.004, 0.015)
σ^2			2.266	(2.237, 2.297)
ϕ			0.718	(0.704, 0.731)
τ^2	2.10	(2.08, 2.13)	0.050	(0.048, 0.053)

Metric	Out-of-sample	In-sample	Out-of-sample	In-sample
DIC	17'588'058		973'329	
Coverage	0.95	0.95	0.93	0.99
RMSPE (r)	1.44 (0.68)	1.44 (0.68)	0.55 (0.09)	0.1 (0.003)
PIW	5.69	5.69	2	1.18

TABLE 2

Parameter estimates and model performance metrics for model (8) with and without the temporal process.

ately around and in the predefined Westwood/UCLA study area (such as the south central zone) or Century City shopping center (to the east). Lighter shades (orange) correspond to areas that are less developed (open space), such as the areas in the north east; or they are areas with a high degree of transportation infrastructure and traffic (e.g., toward the western boundary). These correspond to highways (such as the Interstate-405 highway or other vehicular transportation corridors) that often have lower levels of activity because they inhibit outdoor physical activities due to noise, pollution, safety, etc. Our analysis reveals three additional high activity areas that are not gleaned from non-spatial models: the Los Angeles National Veteran Park; the Century City shopping center and the Stone Canyon Park. The color gradient closely follows the spatial characteristics of the Westwood neighborhood and reveal how spatial patterns can impact physical activity behavior after accounting for variation attributable to known explanatory variables.

Figure 8 shows two examples of observed (left) and reconstructed (right) MAGs along trajectories carved out by two subjects. We find a good degree of agreement between the two plots, and the ability of our model to recover the $\log(\text{MAG})$ in locations where it has not been observed. The reliability of the predictions can be demonstrated through different metrics and, unsurprisingly, accommodating spatial effects and the temporal process improves predictive performances as measured by MSPE or PIW. We deliver these personalized trajectory plots for every subject in the study and also predict personalized MAGs for each subject along any new trajectory. This enables personalized recommendations based upon an individual's health attributes including suggestions for more effective paths to follow for optimal physical activities, while also informing community level interventions in the built environment.

5. Discussion. We have devised a Bayesian modeling framework to conduct fully model-based inference for high-resolution accelerometer data over trajectories compiled from the PASTA-LA study. Our key data analytic developments included (i) modeling dependence over trajectories; (ii) accounting for subject-specific spatial-temporal variation for
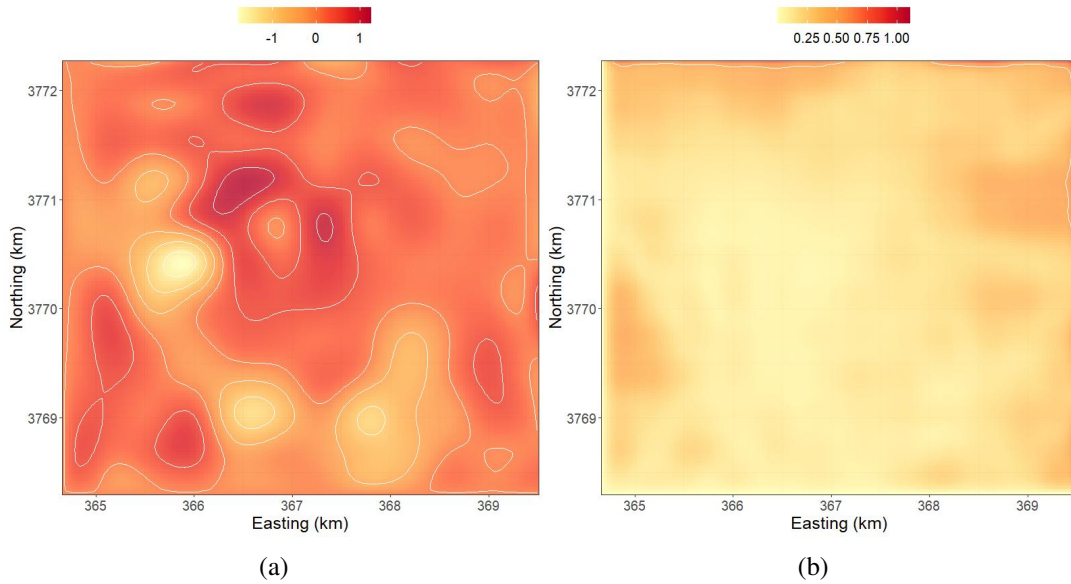
Fig 7: (a) Spatially smoothed estimates from a shrinkage spline over Westwood, Los Angeles; (b) Standard deviation for the shrinkage spline.

daily mobility; and (iii) predicting or interpolating PA levels across trajectories; and (iv) identify zones of high physical activity in Westwood, Los Angeles. Our spatiotemporal analysis offers richer inference and reveals relationships between physical activity levels and a variety of factors, both at the subject level (e.g., personal attributes) and as a function of space and time. The temporal process was able to effectively extract the features of the data at finer resolutions, while the spatial splines accounted for residual spatial heterogeneity. Accommodating both temporal dependence and spatial heterogeneity demonstrably improved predictive ability and enabled us to effectively delineate zones of high physical activity. Furthermore, the ability of the model to pool information across individuals at all time points allows us to infer about those who present sparsely observed space-time points (due to technical issues or protocol violation). In particular, we can interpolate and infer about PA levels with full uncertainty quantification and ensure the desired coverage by our prediction intervals. The methods we develop can be adapted to model animal tracking and be compared to existing spatial models (see, e.g., Hedley and Buckland, 2004).

Recent public health reviews call for interdisciplinary technological advances to more effectively measure spatiotemporal energetics of activity spaces in obesity and chronic disease research (James et al., 2016; Kestens et al., 2017; Drewnowski et al., 2020). Individual-level data, at aggregate, can be used to identify anchor points for physical activity and reveal causal pathways between built environment exposures and health. Our work is a novel contribution demonstrating methodologies to answer these pressing research questions.

Our analysis also resolves practical difficulties in using actigraph data. It is not cost-effective to deploy research-grade GlobalSat GPS and Actigraph units as they are very expensive and continued usage requires heavy staff involvement. Our methods can be applied to analogous, but less complete, data derived from smart phones and smart watches, then such devices could be deployed in much larger studies with much larger sample sizes at a fraction of the cost. Given the spatiotemporal nature of outdoor PA research, our ability to predict in areas of data missingness drastically improve inference related to the impacts of the built and natural environments on physical activity and active mobility.
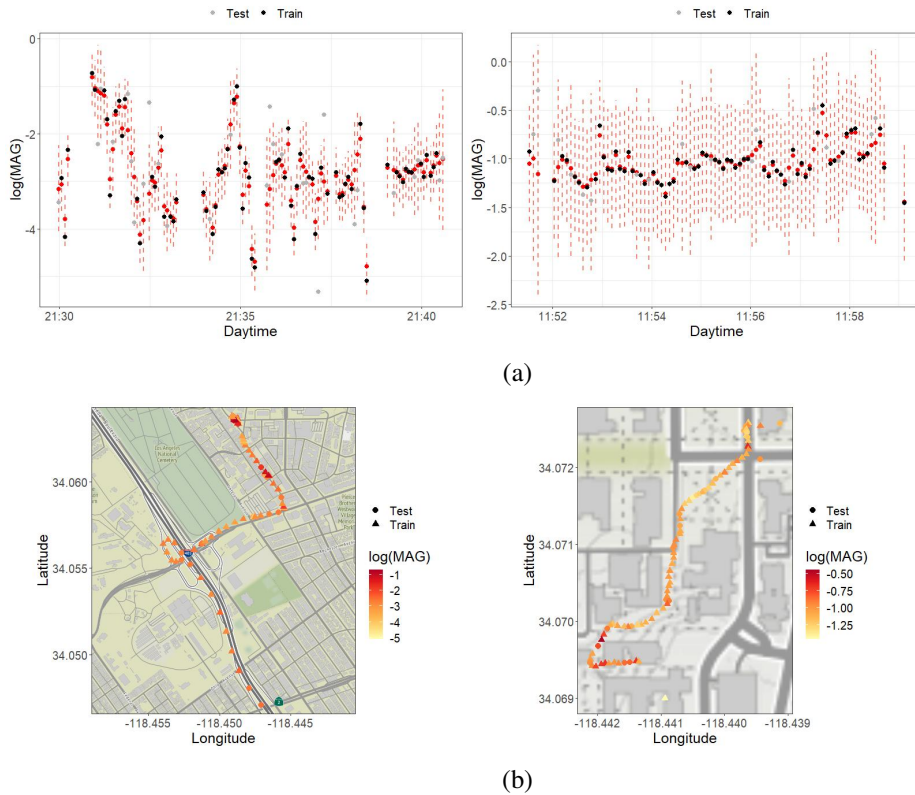
(a)

(b)

Fig 8: Estimated log(MAG) for two randomly selected individuals: (a) estimated log(MAG) (red points) and 95% prediction intervals (red dashed line) for each point within the observed time-windows; (b) including the location.

We recognize that there are several avenues for further research. Substantive investigations pertaining to the PASTA-LA study will focus on the impact of intervention schemes designed to promote physical activities and ask questions related to controlling for weather while estimating the impact of the intervention. Our DAG-based approach for scalable temporal processes can be further enriched with recent developments (Katzfuss and Guinness, 2021; Peruzzi, Banerjee and Finley, 2022), although any of the methods reviewed and evaluated by Heaton et al. (2019) can be incorporated into our framework. We also recognize a wealth of future research surrounding wearable devices and actigraphy data. Examples include methodological advancements in clustering of trajectories according to different levels of physical activity and creating personalized health recommendation systems for patients with regard to trajectories (e.g., walking, running or biking routes) that will be most appropriate for them. Related to the clustering of trajectories, one can also pursue model-based learning about individual effects from the extent of (appropriately quantifies) spatial overlap in trajectories and discerning them from spatial effects Finally, there is possible merit in modeling both the non-idle and idle times with a more comprehensive hierarchical specification (Bai et al., 2018). The joint modeling could be achieved using Mixture Models, Hidden Markov Models, or the modeling of multivariate Gaussian censored outcomes (De Oliveira, 2005; Molstad, Hsu and Sun, 2021). Combining such approaches with efficient estimation strategies is the major challenge, which will be tackled in future developments of this work.

SUPPLEMENTARY MATERIAL

6. Details about data processing. Figure 9 depicts a snapshot of the raw acceleration recorded by the three axles subjected to digital filtering executed by the band-pass Butterworth filter of order 4 in the frequency window $(0.25, 10)$. The elimination of low frequency bands makes the signals stationary around zero, canceling protracted times away from the origin. The effect of the high frequency filtering is less evident, but ensures the cleaning of (typically) low signals occurring at very high frequency.

Figure 10 presents a snapshot of the effect on the point-wise MAG evaluated from the filtered acceleration signals when it is smoothed through the Gaussian kernel according to Equation (2). Significant, but extemporaneous, spikes which are not really indicative of the overall body movement are tampered. The intensity is, instead, quantified as a weighted average of all the neighboring signals. Indeed, the body movement is relevant to PA only if it is protracted for a sufficiently large time window. The single instant, which is a null-measure set, is not as informative as the local average intensity.
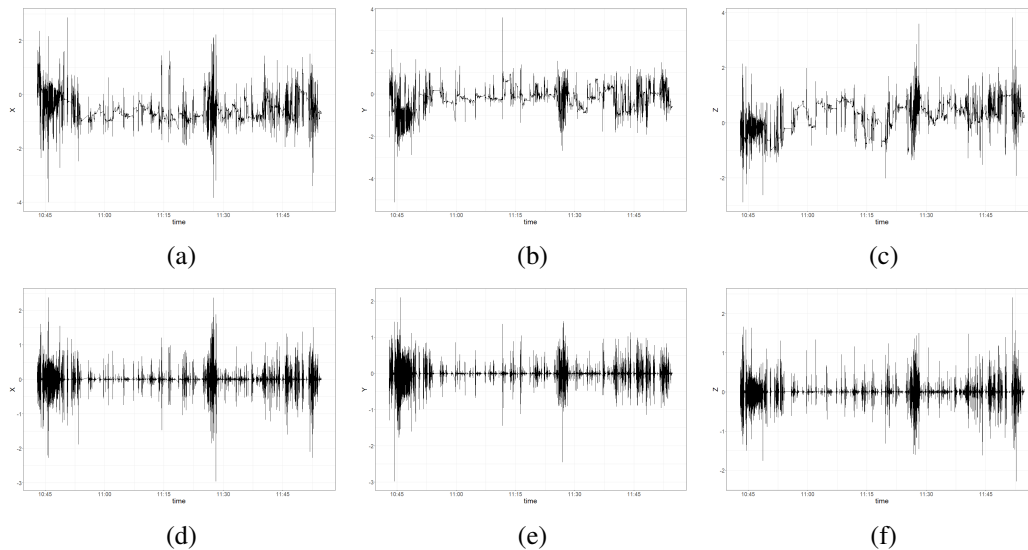

Fig 9: Snapshot of the pre (a-c) and post (d-f) filtering of the axles of a single individual.
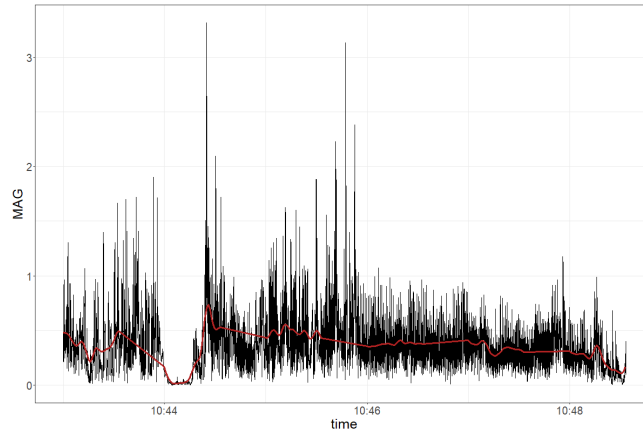
Fig 10: Snapshot of observed (black) and smoothed (red) MAG for a single individual.

7. Relationship between vector magnitude and MET. An extensive review of proposed accelerometer measurement cut-points and transformation into physical activity metrics is available in Migueles et al. (2017). Most of the studies are performed on specific populations, in controlled environments, or during controlled activities. Results are not always coherent with one another based upon current technological capabilities, which has been stoking a significant amount of research. Ellis et al. (2014), Hildebrand et al. (2014), and Staudenmayer et al. (2015) studied the relationship between the raw acceleration MAG, as derived from the GT3X+, and various PA measures using different regression techniques. They found good correlation in both controlled and free-living environments. In particular, Staudenmayer et al. (2015) validated the results with the EE and METs as quantified by a portable calorimeter and compared the accuracy in predicting them when the device is worn on different parts of the body. The author(s) found the dominant wrist to be as good as any other position when the MAG is derived from the raw accelerations instead of the Actilife proprietary software.

While Ellis et al. (2014) uses a Random Forest algorithm and Staudenmayer et al. (2015) considers multiple summary measures of the MAG signal to estimate the corresponding PA, Hildebrand et al. (2014) provides a linear function that relates MAG with VO2 which, in turn, correspond to MET cutpoints. The MAG-to-VO2 relationship for adults wearing an Actigraph on the wrist expounded in Hildebrand et al. (2014) is:

$$(11) \qquad \mathrm{VO2}_{kt} = 0.032 \cdot \mathrm{MAG}_{kt} + 7.28 \,,$$

where the MAG is expressed in mG and averaged per minute. The corresponding MET cutpoints to classify different PA intensity level are in Table 3.

Activity intensity	MET range	MAG (mG)
Sedentary or light	$[0, 3)$	$[0, 100.6)$
Moderate	$[3, 6)$	$[100.6, 428.8)$
Vigorous	$[6, \infty)$	$[428.8, \infty)$

TABLE 3
MAG activity count cut-points for different PA intensity levels.

Based upon the aforementioned literature, inference for the MAG is transformed into METs through (11) to interpret results from a physical activity perspective. Nevertheless,

equations directly relating accelerometer measurements with physical activity metrics in free-living studies must be interpreted with caution. Relationships between MAG and MET have been posited in controlled studies and validated while patients are performing specific tasks (i.e. walking on a treadmill, gardening etc.). The relationship between the recorded movement (acceleration) and the corresponding energy expenditure can vary significantly across different tasks affecting the reliability of acceleration-based energy expenditure metrics (Lyden et al., 2011; Freedson et al., 2012; Montoye et al., 2018; Migueles et al., 2019).

8. Technical details about the implementation of the Collapsed Temporal NNGP. Algorithms 1 and 2 provide the steps for estimating the proposed model. The accompanying website, `https://github.com/minmar94/EfficientTNNGPforActigraph`, supplies the computer codes to fit the proposed model with some examples on simulated data. Algorithm 1 follows the developments in Finley et al. (2019) with two modifications. First, we exploit the temporal domain to make it even more efficient in the neighbor search and in avoiding the fill-in permutation matrix. Second, we generalize to deal with a block-diagonal covariance matrix to account for multiple independent individuals with nested dependent structures. Algorithm 2 is invoked in Step 2 of Algorithm 1 as it replaces typical Bayesian regression updates with the penalized approach developed in Section 3.4.

Algorithm 1: Sampling from the posterior of the collapsed temporal model

0: Initialization
begin
    for $k = 1, \ldots, K$ **do**
        a: Compute $d_{ij}^k = |t_j - t_i|, \quad \forall t_j, t_i \in \mathcal{T}_k$
        b: Find the neighbor sets $\{N_k(i)\}_{i=1}^{T_k}$
    end
end

1: Metropolis-Hastings update for $\{\boldsymbol{\theta}, \tau^2\}$
$p(\boldsymbol{\theta}, \tau^2 | \cdot) \propto p(\boldsymbol{\theta}, \tau^2) \times \frac{1}{\sqrt{\det \widetilde{\boldsymbol{\Lambda}}}} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \widetilde{\boldsymbol{\Lambda}}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right)$
begin
    for $k = 1, \ldots, K$ **do**
        a: Compute $\boldsymbol{L}_k = \left(\boldsymbol{I}_{T_k} - \boldsymbol{A}_k\right)^\top$, $\boldsymbol{d}_k = \mathrm{diag}(\boldsymbol{D}_k)$ and $\boldsymbol{R}_k = \boldsymbol{D}_k^{-1}\left(\boldsymbol{I}_{T_k} - \boldsymbol{A}_k\right)$ using $\boldsymbol{C}_k$ and $\{N_k(i)\}_{i=1}^{T_k}$
        b: Compute $\boldsymbol{\Omega}_k = \boldsymbol{L}_k \cdot \boldsymbol{R}_k + \tau^{-2}\boldsymbol{I}_{T_k}$ exploiting sparsity
        c: Compute $\boldsymbol{r}_k = \boldsymbol{y}_{k\cdot} - \boldsymbol{X}_{k\cdot}\boldsymbol{\beta}$ and $\delta_{\boldsymbol{D}_k} = \prod_{i=1}^{T_k} d_{k,i}$
        d: Compute $\boldsymbol{v}_k = \boldsymbol{\Omega}_k^{-1}\boldsymbol{r}_k$, $\boldsymbol{u}_k = \boldsymbol{\Omega}_k^{-1}\boldsymbol{X}_{k\cdot}$ and $\delta_{\boldsymbol{\Omega}_k} = \det(\boldsymbol{\Omega}_k)$ exploiting the sparse Cholesky decomposition of $\boldsymbol{\Omega}_k$
        e: Collect $\boldsymbol{r}_k$, $\boldsymbol{v}_k$ and $\boldsymbol{u}_k$ into $\boldsymbol{r}$, $\boldsymbol{v}$ and $\boldsymbol{u}$, respectively.
    end
    f: Compute $q_1 = \tau^{2n} \cdot \prod_{k=1}^K \delta_{\boldsymbol{D}_k} \cdot \prod_{k=1}^K \delta_{\boldsymbol{\Omega}_k}$ and $q_2 = \boldsymbol{r}^\top \boldsymbol{r}/\tau^2 - \boldsymbol{r}^\top \boldsymbol{v}/\tau^4$
    g: Get $p(\boldsymbol{\theta}, \tau^2 | \cdot) \propto \frac{\exp(-q_2/2)}{\sqrt{q_1}} \cdot p(\boldsymbol{\theta}, \tau^2)$
end

2: Gibbs' sampler update for $\boldsymbol{\beta}$
$\boldsymbol{\beta} | \cdot \sim \mathcal{N}_p(\boldsymbol{B}^{-1}\boldsymbol{b}, \boldsymbol{B}^{-1})$, where $\boldsymbol{B} = \boldsymbol{X}^\top \widetilde{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X} + \boldsymbol{V}_{\boldsymbol{\beta}}^{-1}$ and $\boldsymbol{b} = \boldsymbol{X}^\top \widetilde{\boldsymbol{\Lambda}}^{-1}\boldsymbol{y} + \boldsymbol{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}}$
begin
    a: Compute $\boldsymbol{F} = \boldsymbol{V}_{\boldsymbol{\beta}}^{-1}$ and $= \boldsymbol{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}}$
    b: Compute $\boldsymbol{b} = \boldsymbol{y}^\top \boldsymbol{X}/\tau^2 - \boldsymbol{y}^\top \boldsymbol{v}/\tau^4 +$ and $\boldsymbol{B} = \boldsymbol{X}^\top \boldsymbol{X}/\tau^2 - \boldsymbol{X}^\top \boldsymbol{v}/\tau^4 + \boldsymbol{F}$
    c: Generate $\boldsymbol{\beta} \sim \mathcal{N}_p(\boldsymbol{B}^{-1}\boldsymbol{b}, \boldsymbol{B}^{-1})$
end
Repeat steps **1** and **2** to obtain M MCMC samples for $\{\boldsymbol{\beta}, \boldsymbol{\theta}, \tau^2\}$

Algorithm 2: ψ and λ Gibbs' update in the collapsed algorithm with shrinkage

1: Gibbs' sampler update for ψ

$\psi|\cdot \sim \mathcal{N}_J(\boldsymbol{G}^{-1}\boldsymbol{g}, \boldsymbol{G}^{-1})$, where $\boldsymbol{G} = \boldsymbol{X}^{*\top}\widetilde{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}^* + \boldsymbol{V}_{\psi}^{-1}$ and $\boldsymbol{g} = \boldsymbol{X}^{*\top}\widetilde{\boldsymbol{\Lambda}}^{-1}\boldsymbol{y} + \boldsymbol{V}_{\psi}^{-1}\boldsymbol{\mu}_{\psi}$

begin

    a: Compute $\boldsymbol{F} = \boldsymbol{V}_{\psi}^{-1}$ and $= \boldsymbol{V}_{\psi}^{-1}\boldsymbol{\mu}_{\psi}$

    b: Compute $\boldsymbol{g} = \boldsymbol{y}^{\top}\boldsymbol{X}^*/\tau^2 - \boldsymbol{y}^{\top}\boldsymbol{v}/\tau^4 +$ and $\boldsymbol{G} = \boldsymbol{X}^{*\top}\boldsymbol{X}^*/\tau^2 - \boldsymbol{X}^{*\top}\boldsymbol{v}/\tau^4 + \boldsymbol{F}$

    c: Generate $\psi \sim \mathcal{N}_{p^*}(\boldsymbol{G}^{-1}\boldsymbol{g}, \boldsymbol{G}^{-1})$

end

2: Gibbs' sampler update for λ

$\lambda|\cdot \sim Ga\left(\alpha_{\lambda}^*, \beta_{\lambda}^*\right)$, where $\alpha_{\lambda}^* = \alpha_{\lambda} + 1/2$ and $\beta_{\lambda}^* = \beta_{\lambda} + \boldsymbol{\beta}_S^{\top}\boldsymbol{P}\boldsymbol{\beta}_S$

begin

    a: Compute $h = \boldsymbol{\beta}_S^{\top}\boldsymbol{P}\boldsymbol{\beta}_S$ and get: $\alpha_{\lambda}^* = \alpha_{\lambda} + 1/2$ and $\beta_{\lambda}^* = \beta_{\lambda} + h$

    b: Generate $\lambda \sim \mathcal{G}\left(\alpha_{\lambda}^*, \beta_{\lambda}^*\right)$

end

---

9. Additional simulation experiments. We carried out two additional experiments to test the reliability of our algorithm and verify comparative performances with the "Sequential NNGP" as it is implemented in the `spNNGP` package (Finley, Datta and Banerjee, 2017). We did not consider the Response NNGP because it is not designed to recover the latent component. Our first experiment is described in Section 9.1 and includes simulated observations for one single individual; the second one includes simulated observations for multiple individuals and is described in Section 9.2. Codes to reproduce the following results and additional comparative analyses of NNGP versus the full GP model are available at https://github.com/minmar94/EfficientTNNGPforActigraph.

9.1. *Experiment 1.* We generated observations $\{y(t_j)\}_{j=1}^{\top}$ for $K = 1$ individual, using $T = 10^5$ time-points, where each $t_i = \sum_{h=1}^{i-1}\delta_h$, and $\delta_h \sim Exp(5)$, $\forall h$. The model included an intercept β_0 and 3 covariates, x_1, x_2 and x_3 all drawn from $\mathcal{N}(0,1)$, with associated slopes β_1, β_2 and β_3. We modeled the covariance structure between any two simulations at time-points t and t' using the exponential covariance function,

$$(12) \qquad \text{Cov}_{\theta}\left[Y(t), Y(t')\right] = c_{\theta}(t, t') = \sigma^2 e^{(\phi|t-t'|)}, \quad \sigma^2, \phi \in \mathbb{R}^+,$$

where σ^2 represents the variance of the process (sill), ϕ is the decay in temporal correlation (inversely proportional to range) and τ^2 the residual variance (nugget). In this data generation step the parameters have been set to the following values: $\beta_0 = -1.878$, $\beta_1 = 0.326$, $\beta_2 = -0.302$, $\beta_3 = 1.182$, $\sigma^2 = \phi = \tau^2 = 1$. A chunk of the simulated trajectory and its density can be observed as an example in Figures 11a and 11b, respectively.

We fitted the model on the simulated data using our *Collapsed NNGP* implementation, specifically optimized for the temporal setting, while fitting the *Sequential* NNGP using the `spNNGP` package. The latter, while generally used for fitting spatial (i.e. two-dimensionals) models, can be adapted to the temporal (uni-dimensional) case by providing a set of locations where t is one of the coordinates and the other is fixed to a constant value (e.g. $\{\widetilde{\boldsymbol{s}}_j\}_{j=1}^{\top} = \{(t_j, 0)\}_{j=1}^{\top}$). In our implementation, the intercept and slope were assigned vague normal prior distributions ($\mathcal{N}(0, 10^6)$). The variance components, σ^2 and τ^2, were both assigned inverse Gamma priors ($\mathcal{IG}(2, 2)$) and the decay parameter, ϕ, was assigned a Gamma prior $\mathcal{G}(1, 1)$.

On the other hand, `spNNGP` assumes a flat prior on the intercept and slope coefficients and a uniforma $\mathcal{U}(a, b)$ prior on the decay parameter ϕ. In this experiment we fixed $a = 0.5$ and
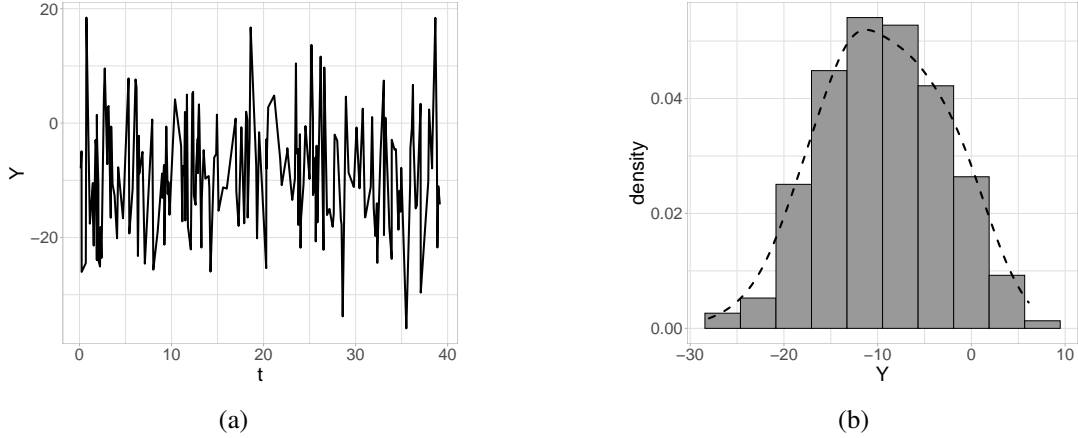
Fig 11: Simulated uni-dimensional Gaussian process (a) and its density (b).

Param. (True)	Collapsed NNGP			Sequential NNGP		
	Point	Interval	ESS	Point	Interval	ESS
$\beta_0\,(-1.88)$	-1.87	(-1.89, -1.85)	4999	-1.87	(-1.89, -1.85)	57
$\beta_1\,(0.33)$	0.33	(0.32, 0.34)	4999	0.33	(0.32, 0.34)	1285
$\beta_2\,(-0.30)$	-0.30	(-0.31, -0.29)	4999	-0.30	(-0.31, -0.3)	1365
$\beta_3\,(1.18)$	1.18	(1.17, 1.19)	4999	1.18	(1.17, 1.19)	1342
$\sigma^2\,(1)$	1.00	(0.97, 1.03)	472	1.00	(0.97, 1.03)	294
$\phi\,(1)$	0.99	(0.95, 1.04)	496	0.99	(0.95, 1.04)	65
$\tau^2\,(1)$	1.01	(0.99, 1.03)	457	1.01	(0.99, 1.03)	165
Metric	Out-of-sample	In-sample		Out-of-sample	In-sample	
Coverage	0.95	0.99		0.96	0.99	
RMSPE (r)	0.39 (1.19)	0.20 (0.85)		0.39 (1.19)	0.20 (0.85)	
PIW	4.68	4.46		4.78	4.47	
Run time (h)		1.77			1.86	

TABLE 4
Parameter estimates, predictive validation and fitting times (hours) on the simulated dataset for all the considered models.

$b = 30$. All the models were trained on the same random sample composed of 70% of the total observations, while the the remaining 30% were held out to evaluate the out-of-sample predictive performances using *Relative Mean Squared Prediction Error* (RMSPE), *Root Mean Squared Prediction Error* (rMSPE), *Coverage*, and *Predictive Interval Width* (PIW).

We fixed the number of neighbors $m = 10$ in the DAG and executed $10,000$ MCMC iterations. The first $5,000$ simulations were dropped as burn-in, while the last $5,000$ were retained for estimation and prediction purposes. No thinning has been considered. Results are summarized in Table 4. The two approaches provide substantively indistinguishable inference in terms of estimation and prediction. However, our implementation is faster than its competitor (at least in the context of the temporal setting) and provides pronouncedly better performances in terms of Effective Sample Size (ESS).

9.1.1. *Computation time evaluation.* We *quantified* the *linearity* of all the algorithms: by construction, the fitting time should increase linearly with the sample size. We split observations in $l = 1, \dots, 5$ different fitting windows $\{t_1, \dots, t_{T_l}\}$ with increasing sizes $T_l = \{\{2^l\}_{l=0}^6 \cup \{100\}\} \times 10^3$. The computation time for each iteration, fixing $m = 30$
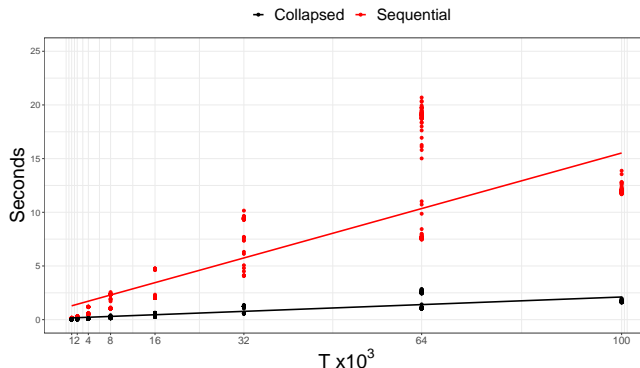
Fig 12: Time elapsed (in seconds) for 1 MCMC iteration for the two considered algorithms with increasing sample size T and fixed $m = 30$.

neighbors in the DAG, for all the considered algorithms has been recorded with $\widetilde{M} = 100$. Figure 12 shows that all algorithms scale linearly with the sample size. However, our implementation of the collapsed NNGP, while pointed out as generally less efficient than its competitors in Finley et al. (2019), scales with a rate of $\approx 0.376 \cdot 10^{-4}$ per data point, while the Sequential NNGP scales with a rate equal to $\approx 4.5736 \cdot 10^{-4}$, which is ostensibly higher. Table 5 presents detailed results.

$T \times 10^3$	Algorithm	Min	q_{025}	Median	Mean	q_{975}	Max
1	Collapsed	0.01	0.01	0.02	0.02	0.03	0.03
	Sequential	0.12	0.12	0.13	0.14	0.18	0.18
2	Collapsed	0.03	0.03	0.03	0.04	0.06	0.07
	Sequential	0.25	0.25	0.26	0.27	0.31	0.34
4	Collapsed	0.06	0.06	0.07	0.09	0.16	0.16
	Sequential	0.50	0.50	0.52	0.64	1.21	1.21
8	Collapsed	0.13	0.14	0.30	0.26	0.32	0.41
	Sequential	1.01	1.01	2.34	1.99	2.41	2.56
16	Collapsed	0.27	0.28	0.60	0.46	0.63	0.64
	Sequential	2.02	2.02	4.65	3.46	4.75	4.77
32	Collapsed	0.55	0.56	1.23	1.17	1.28	1.37
	Sequential	4.08	4.09	9.40	8.87	9.59	10.16
64	Collapsed	1.01	1.03	2.46	1.90	2.79	2.85
	Sequential	2.51	7.49	18.71	14.43	20.13	20.69
100	Collapsed	1.60	1.61	1.67	1.68	1.87	1.99
	Sequential	11.68	11.74	11.93	12.01	12.80	13.87

TABLE 5

Time (in seconds) of one MCMC iteration for the two considered algorithms with increasing sample size (T) and fixed $m = 30$.

Additionally, we wanted to quantify the computational advantage of the proposed NNGP-based collapsed algorithm over the standard MCMC update of the full GP model, where the latter is already implemented in R through the spLM() function in the spBayes package (Finley, Banerjee and Gelfand, 2013). Since Datta et al. (2016a) demonstrated that the NNGP approximation with $m = 30$ neighbors provides almost exactly the same inference of the full GP, we fixed $m = 30$ and built different sets of data with increasing number of data-points,
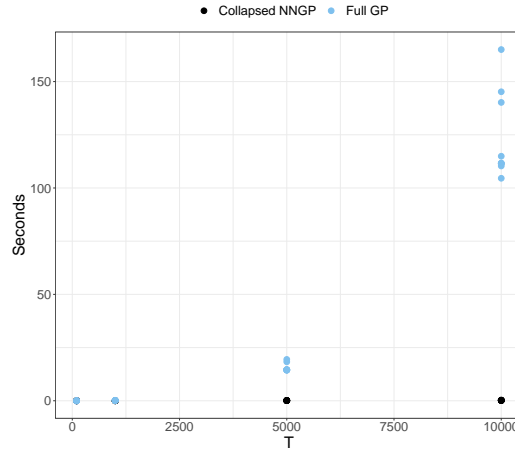
Fig 13: Time elapsed (in seconds) for 1 MCMC iteration for the Collapsed NNGP and the Full GP with increasing sample size T.

but this time with $T_l = 100, 1,000, 5,000, 10,000$ (sizes have been reduced to accommodate the slow update of the full GP). Figure 13 shows that for $T_l = 100$ the computational time difference between the full GP and the collapsed NNGP is negligible. However, as the size increases the savings increase exponentially: 15 seconds (per iteration) when $n = 5,000$ and 122 seconds per iteration when $n = 10,000$. For the last scenario, considering 10,000 MCMC iterations, the collapsed NNGP will provide results 14 days in advance of the Full GP model.

9.2. *Experiment 2.* The aim of this experiment is to verify the ability of our algorithm in recovering the true parameters and to determine if pooling information from multiple individuals can help in improving the accuracy of the estimates. Comparisons with the *Sequential NNGP* is not feasible because the latter precludes the contemporary fitting of multiple Gaussian processes with common parameters. Thus, we compare performances of the Pooled NNGP (that's how we will refer to the collapsed algorithm in what follows) with the single models estimated separately for each individual.

We generated $2 \cdot 10^4$ observations for $K = 5$ individuals, using the same scheme as in Experiment 1 (total of 10^5 data-points). Results are presented in Table 6. The model also included 3 covariates and an intercept for each individual drawn from independent $N(0,1)$. Observations were then generated as described in Section 9.1. The simulated data was split into two sets: 70% constituted the training set for estimation purposes, while the remaining 30% was used to assess model predictive performances. RMSPE, coverage of the predictive 95% credible intervals and their mean width were used as measures of the goodness of fit.

For all the models, the intercept and slope parameters were assigned a flat normal prior distribution $\mathcal{N}(0, 10^6)$. The variance components, σ^2 and τ^2, were both assigned inverse Gamma $\mathcal{IG}(2,2)$ priors, while the decay parameter ϕ received a Gamma prior $\mathcal{G}(1,1)$. The advantage of pooling information from multiple individuals for the estimation of common parameters, while the independence assumption among them still holds, is evident according to all criteria. First, there is a sensible gain in estimating the accuracy of the common parameters. Indeed, while the true value of the parameters are included in the intervals also considering one single individual at a time, the widths of 95% credible intervals are sensibly smaller when we pool the information together. Furthermore, some slight advantage is also visible for prediction purposes, where the Pooled NNGP delivers larger coverage and smaller

Param.	True	Pooled NNGP	Individuals 1	2	3	4	5
β_{01}	-9.39	-9.41 (-9.46, -9.37)	-9.41 (-9.46, -9.37)				
β_{02}	1.63	1.59 (1.54, 1.64)		1.59 (1.54, 1.64)			
β_{03}	-1.51	-1.53 (-1.57, -1.48)			-1.54 (-1.58, -1.49)		
β_{04}	5.91	5.91 (5.86, 5.96)				5.91 (5.86, 5.96)	
β_{05}	-0.82	-0.80 (-0.85, -0.76)					-0.81 (-0.86, -0.76)
β_1	6.48	6.48 (6.47, 6.49)	6.48(6.46,6.5)	6.48(6.46,6.50)	6.48(6.46,6.50)	6.46(6.44,6.48)	6.49(6.47,6.51)
β_2	6.76	6.75 (6.74, 6.76)	6.74(6.72,6.76)	6.75(6.74,6.77)	6.75(6.73,6.77)	6.75(6.73,6.77)	6.76(6.74,6.78)
β_3	-1.46	-1.46 (-1.47, -1.45)	-1.47(-1.49,-1.45)	-1.48(-1.50,-1.46)	-1.47(-1.48,-1.45)	-1.45(-1.47,-1.43)	-1.47(-1.49,-1.45)
σ^2	1	0.98 (0.96, 1.01)	0.94(0.88,1.00)	1.02(0.953,1.085)	0.93(0.88,0.99)	1.03(0.97,1.1)	1.01(0.94,1.08)
ϕ	1	1.01 (0.97, 1.06)	1.00(0.91,1.11)	1.03(0.93,1.14)	1.06(0.95,1.17)	0.94(0.85,1.04)	0.99(0.9,1.11)
τ^2	1	1 (0.99, 1.02)	1.04(1.00,1.08)	0.98(0.94,1.02)	1.00(0.967,1.04)	1.00(0.96,1.04)	0.99(0.95,1.03)
Coverage		0.95 (0.99)	0.95 (0.99)	0.95 (0.99)	0.95 (0.99)	0.96 (0.99)	0.95 (0.99)
RMSPE		0.012 (0.006)	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)
rMSPE		1.22 (0.84)	1.24 (0.83)	1.24 (0.83)	1.22 (0.85)	1.22 (0.85)	1.23 (0.84)
PIW		4.67 (4.44)	4.99 (4.44)	4.94 (4.43)	4.93 (4.43)	4.95 (4.44)	4.94 (4.44)
Fitting time		0.59	0.33	0.30	0.35	0.34	0.37

Table 6: Parameter estimates (credible intervals ($q_{0.025}$, $q_{0.975}$)), *out-of-sample* prediction error and fitting times (hours) on the simulated dataset for the pooled and the single models.

RMSPE. Additionally, due to the parallel implementation of our code, there is almost no loss in terms of the computational time required for the fitting: ≈ 40 minutes to fit one individual as opposed to ≈ 55 minutes to fit the pooled model.
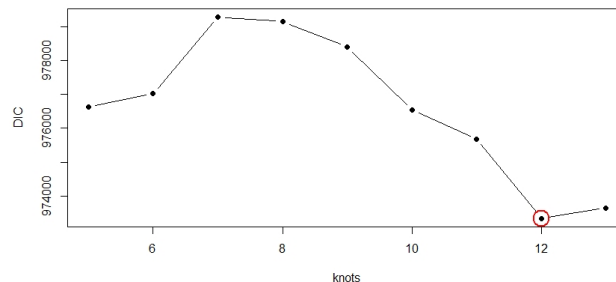

Fig 14: Selection of the optimal number of knots for the basis of the B-splines.

10. Further details about the results in Section 4. This section provides a few additional details about the estimates resulting from Section 4. Figure 14 shows the behavior of the DIC as a function of the number of knots for the basis of the B-splines accounting for the non-parametric spatial effect. We note that the DIC starts improving when the number of knots is greater than 7 and proceeds to offer the best fit when the number of knots is equal to 12. Figure 15 contains some basic residual diagnostics aimed at verifying if the performances are stable across the whole outcome domain and also indicates the Gaussianity of the residuals. Figure 16 provides the traceplot of the chains of 6 parameters to check their convergence and stationarity. We report two randomly selected β's, the shrinkage parameter λ, and ϕ, σ^2, τ^2. All other β's show comparable patterns. Figure 17 illustrates the capability of the estimated model (see Section 4) to approximate the marginal distribution of the observed outcome, although it is bimodal. This is not surprising because the temporal Gaussian process acts as a non-parametric prior that interpolates the observed data (see, e.g., Banerjee, 2019), which makes it a highly effective model for high-resolution data even if the distribution of the observations is decidedly non-Gaussian.

REFERENCES

BAI, J., SUN, Y., SCHRACK, J. A., CRAINICEANU, C. M. and WANG, M.-C. (2018). A two-stage model for wearable device data. *Biometrics* **74** 744–752. 2, 19

BAMMANN, K., THOMSON, N. K., ALBRECHT, B. M., BUCHAN, D. S. and EASTON, C. (2021). Generation and validation of ActiGraph GT3X+ accelerometer cut-points for assessing physical activity intensity in older adults. The OUTDOOR ACTIVE validation study. *PloS one* **16** e0252615. 4

BANERJEE, S. (2017). High-Dimensional Bayesian Geostatistics. *Bayesian Analysis* **12** 583-614. 9

BANERJEE, S. (2019). Geostatistical Modeling for Environmental Processes. In *Handbook of Environmental and Ecological Statistics* first ed. (A. E. Gelfand, M. Fuentes, J. Hoeting and R. L. Smith, eds.) 5, 81–95. Chapman and Hall/CRC, Boca Raton, FL. 28

BULL, F. C., AL-ANSARI, S. S., BIDDLE, S., BORODULIN, K., BUMAN, M. P., CARDON, G., CARTY, C., CHAPUT, J.-P., CHASTIN, S., CHOU, R. et al. (2020). World Health Organization 2020 guidelines on physical activity and sedentary behaviour. *British journal of sports medicine* **54** 1451–1462. 1

CRESSIE, N. and WIKLE, C. K. (2015). *Statistics For Spatio-Temporal Data*. John Wiley & Sons. 2

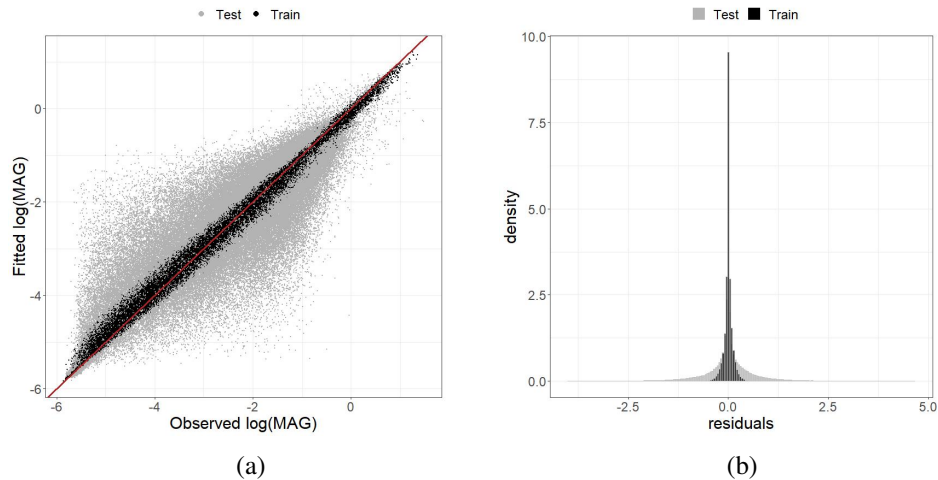(a)   (b)

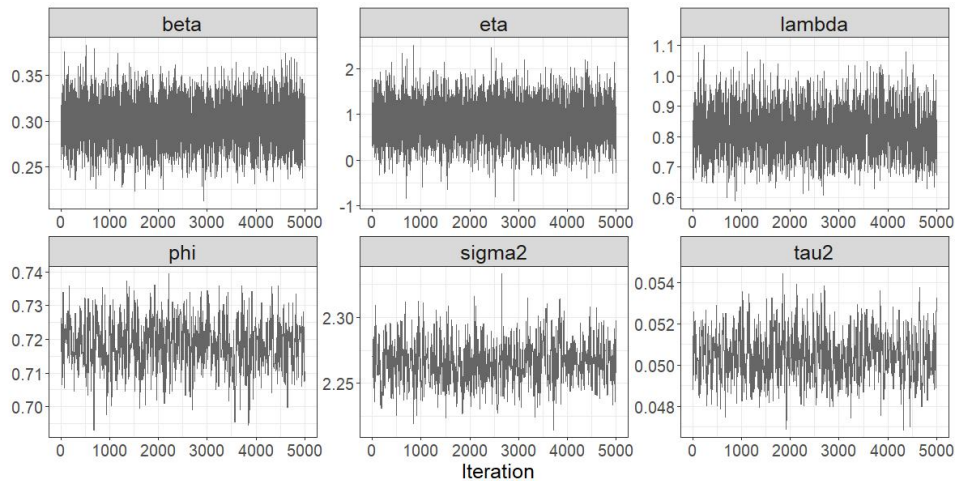Fig 15: Diagnostic check: (a) observed VS fitted values of the log(MAG) on the test set; (b) histogram of the residuals.

Fig 16: Traceplots for two random covariates, the shrinkage parameter λ (top panels), and the covariance parameters ϕ, σ^2, τ^2 (bottom panels).

CROUTER, S. E., CLOWERS, K. G. and BASSETT JR, D. R. (2006). A Novel Method For Using Accelerometer Data To Predict Energy Expenditure. *Journal of applied physiology* **100** 1324–1331. 4

DATTA, A., BANERJEE, S., FINLEY, A. O. and GELFAND, A. E. (2016a). Hierarchical Nearest-Neighbor Gaussian Process Models For Large Geostatistical Datasets. *Journal of the American Statistical Association* **111** 800–812. 9, 10, 25

DATTA, A., BANERJEE, S., FINLEY, A. O., HAMM, N. A. S. and SCHAAP, M. (2016b). Non-Separable Dynamic Nearest-Neighbor Gaussian Process Models For Large Spatio-Temporal Data With An Application To Particulate Matter Analysis. *Annals of Applied Statistics* **10** 1286-1316. 9

DE OLIVEIRA, V. (2005). Bayesian inference and prediction of Gaussian random fields based on censored data. *Journal of Computational and Graphical Statistics* **14** 95–115. 19

DOHERTY, A., JACKSON, D., HAMMERLA, N., PLÖTZ, T., OLIVIER, P., GRANAT, M. H., WHITE, T., VAN HEES, V. T., TRENELL, M. I., OWEN, C. G. et al. (2017). Large Scale Population Assessment Of Physical Activity Using Wrist Worn Accelerometers: The Uk Biobank Study. *PloS one* **12**. 4

DREWNOWSKI, A., BUSZKIEWICZ, J., AGGARWAL, A., ROSE, C., GUPTA, S. and BRADSHAW, A. (2020). Obesity and the Built Environment: A Reappraisal. *Obesity* **28** 22-30. 2, 18
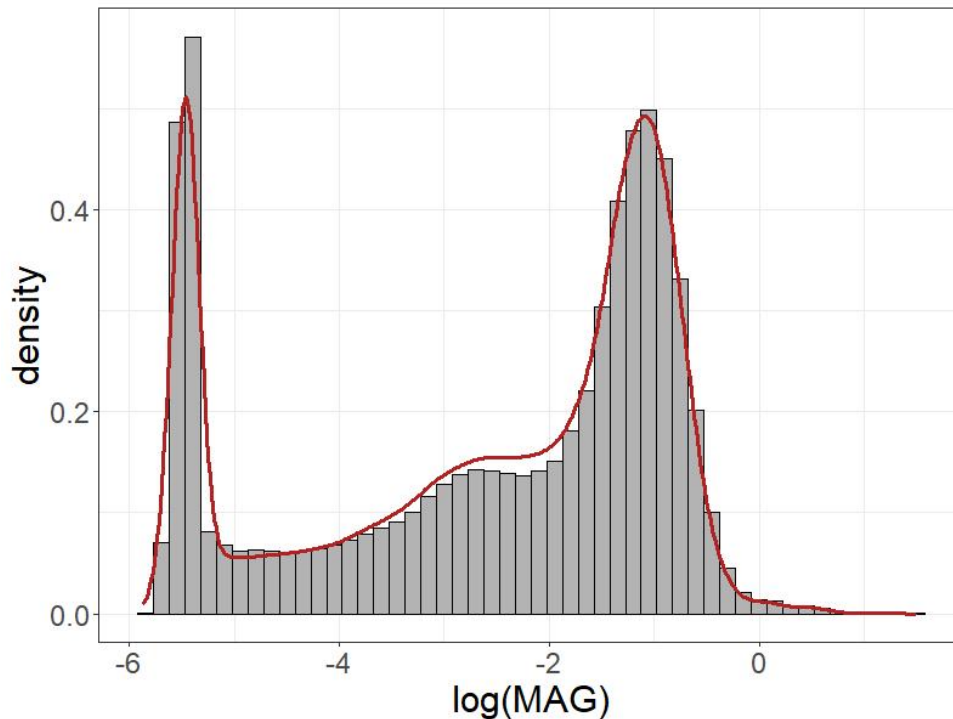
Fig 17: Histogram of the observed values of the log(MAG) and estimated posterior density using model (8) including the temporal process.

EILERS, P. H. and MARX, B. D. (1996). Flexible Smoothing With B-Splines And Penalties. *Statistical science* 89–102. 11

ELLIS, K., KERR, J., GODBOLE, S., LANCKRIET, G., WING, D. and MARSHALL, S. (2014). A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiological measurement* **35** 2191. 21

FINLEY, A. O., BANERJEE, S. and GELFAND, A. E. (2013). spBayes for large univariate and multivariate point-referenced spatio-temporal data models. *arXiv preprint arXiv:1310.8192*. 25

FINLEY, A., DATTA, A. and BANERJEE, S. (2017). Spnngp: Spatial Regression Models For Large Datasets Using Nearest Neighbor Gaussian Processes. *R package version 0.1* **1**. 11, 12, 23

FINLEY, A. O., DATTA, A., COOK, B. D., MORTON, D. C., ANDERSEN, H. E. and BANERJEE, S. (2019). Efficient Algorithms For Bayesian Nearest Neighbor Gaussian Processes. *Journal of Computational and Graphical Statistics* 1–14. 10, 22, 25

FREEDSON, P., BOWLES, H. R., TROIANO, R. and HASKELL, W. (2012). Assessment Of Physical Activity Using Wearable Monitors: Recommendations For Monitor Calibration And Use In The Field. *Medicine and science in sports and exercise* **44** S1. 4, 22

GELFAND, A. E., DIGGLE, P., GUTTORP, P. and FUENTES, M. (2010). *Handbook of Spatial Statistics*. CRC press. 2

GILKS, W. R. and ROBERTS, G. O. (1996). Strategies for Improving MCMC. *Markov chain Monte Carlo in practice* **6** 89–114. 15

GOODMAN, T. and HARDIN, D. (2006). Refinable Multivariate Spline Functions. In *Studies in Computational Mathematics*, **12** 55–83. Elsevier. 11

HAARIO, H., SAKSMAN, E., TAMMINEN, J. et al. (2001). An Adaptive Metropolis Algorithm. *Bernoulli* **7** 223–242. 10

HASTIE, T., TIBSHIRANI, R. et al. (2000). Bayesian Backfitting (with comments and a rejoinder by the authors. *Statistical Science* **15** 196–223. 11

HEATON, M. J., DATTA, A., FINLEY, A. O., FURRER, R., GUINNESS, J., GUHANIYOGI, R., GERBER, F., GRAMACY, R. B., HAMMERLING, D., KATZFUSS, M. et al. (2019). A Case Study Competition Among Methods For Analyzing Large Spatial Data. *Journal of Agricultural, Biological and Environmental Statistics* **24** 398–425. 9, 19

HEDLEY, S. L. and BUCKLAND, S. T. (2004). Spatial models for line transect sampling. *Journal of Agricultural, Biological and Environmental Statistics* **9** 181–199. 18

HILDEBRAND, M., VT, V. H., HANSEN, B. H. and EKELUND, U. (2014). Age group comparability of raw accelerometer output from wrist-and hip-worn monitors. *Medicine and science in sports and exercise* **46** 1816–1824. 21

JAMES, P., JANKOWSKA, M., MARX, C., HART, J. E., BERRIGAN, D., KERR, J., HURVITZ, P. M., HIPP, J. A. and LADEN, F. (2016). "Spatial Energetics": Integrating Data From GPS, Accelerometry, and GIS to Address Obesity and Inactivity. *American Journal of Preventive Medicine* **51** 792-800. 2, 18

KATZFUSS, M. and GUINNESS, J. (2021). A General Framework for Vecchia Approximations of Gaussian Processes. *Statist. Sci.* **36** 124–141. 9, 19

KATZFUSS, M., GUINNESS, J., GONG, W. and ZILBER, D. (2020). Vecchia Approximations of Gaussian-process Predictions. *Journal of Agricultural, Biological and Environmental Statistics* **25** 383–414. 9

KESTENS, Y., WASFI, R., NAUD, A. and CHAIX, B. (2017). "Contextualizing Context": Reconciling Environmental Exposures, Social Networks, and Location Preferences in Health Research. *Current Environmental Health Reports* **4** 51-60. 2, 18

KHUSAINOV, R., AZZI, D., ACHUMBA, I. E. and BERSCH, S. D. (2013). Real-time human ambulation, activity, and physiological monitoring: Taxonomy of issues, techniques, applications, challenges and limitations. *Sensors* **13** 12852–12902. 4

LANG, S. and BREZGER, A. (2004). Bayesian P-Splines. *Journal of computational and graphical statistics* **13** 183–212. 11

LAURITZEN, S. L. (1996). *Graphical Models* **17**. Clarendon Press. 9

LIU, J. S., WONG, W. H. and KONG, A. (1994). Covariance Structure Of The Gibbs Sampler With Applications To The Comparisons Of Estimators And Augmentation Schemes. *Biometrika* **81** 27–40. 10

ALAIMO DI LORO, P., MINGIONE, M., LIPSITT, J., BATTEATE, C. M., JERRETT, M. B. and BANERJEE, S. (2023). Supplement to "Bayesian hierarchical modeling and analysis for actigraph data from wearable devices". *Annals of Applied Statistics*. 4, 10, 11, 12, 14, 16

LYDEN, K., KOZEY, S. L., STAUDENMEYER, J. W. and FREEDSON, P. S. (2011). A Comprehensive Evaluation Of Commonly Used Accelerometer Energy Expenditure And Met Prediction Equations. *European journal of applied physiology* **111** 187–201. 22

LYDEN, K., KEADLE, S. K., STAUDENMAYER, J. and FREEDSON, P. S. (2014). A Method To Estimate Free-Living Active And Sedentary Behavior From An Accelerometer. *Medicine and science in sports and exercise* **46** 386. 4

MADDISON, R., HOORN, S. V., JIANG, Y., MHURCHU, C. N., EXETER, D., DOREY, E., BULLEN, C., UTTER, J., SCHAAF, D. and TURLEY, M. (2009). The environment and physical activity: The influence of psychosocial, perceived and built environmental factors. *International Journal of Behavioral Nutrition and Physical Activity* **6:19**. 5

MATHIE, M., COSTER, A., LOVELL, N. and CELLER, B. (2003). Detection of daily physical activities using a triaxial accelerometer. *Medical and Biological Engineering and Computing* **41** 296–301. 4

MIGUELES, J. H., CADENAS-SANCHEZ, C., EKELUND, U., DELISLE NYSTRÖM, C., MORA-GONZALEZ, J., LÖF, M., LABAYEN, I., RUIZ, J. R. and ORTEGA, F. B. (2017). Accelerometer Data Collection And Processing Criteria To Assess Physical Activity And Other Outcomes: A Systematic Review And Practical Considerations. *Sports medicine* **47** 1821–1845. 4, 21

MIGUELES, J. H., CADENAS-SANCHEZ, C., ROWLANDS, A. V., HENRIKSSON, P., SHIROMA, E. J., ACOSTA, F. M., RODRIGUEZ-AYLLON, M., ESTEBAN-CORNEJO, I., PLAZA-FLORIDO, A., GIL-COSANO, J. J. et al. (2019). Comparability of accelerometer signal aggregation metrics across placements and dominant wrist cut points for the assessment of physical activity in adults. *Scientific reports* **9** 1–12. 4, 22

MOLSTAD, A. J., HSU, L. and SUN, W. (2021). Gaussian process regression for survival time prediction with genome-wide gene expression. *Biostatistics* **22** 164–180. 19

MONTOYE, A. H., MOORE, R. W., BOWLES, H. R., KORYCINSKI, R. and PFEIFFER, K. A. (2018). Reporting Accelerometer Methods In Physical Activity Intervention Studies: A Systematic Review And Recommendations For Authors. *British journal of sports medicine* **52** 1507–1516. 22

MURPHY, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT press. 9

PERUZZI, M., BANERJEE, S. and FINLEY, A. O. (2022). Highly Scalable Bayesian Geostatistical Modeling via Meshed Gaussian Processes on Partitioned Domains. *Journal of the American Statistical Association* **117** 969-982. 9, 19

PIERCY, K. L., TROIANO, R. P., BALLARD, R. M., CARLSON, S. A., FULTON, J. E., GALUSKA, D. A., GEORGE, S. M. and OLSON, R. D. (2018). The physical activity guidelines for Americans. *Jama* **320** 2020–2028. 1

PLASQUI, G. and WESTERTERP, K. R. (2007). Physical Activity Assessment With Accelerometers: An Evaluation Against Doubly Labeled Water. *Obesity* **15** 2371–2379. 2

RAMSAY, J. O. and SILVERMAN, B. W. (2007). *Applied Functional Data Analysis: Methods and Case Studies*. Springer. 11

RAY, E. L., SASAKI, J. E., FREEDSON, P. S. and STAUDENMAYER, J. (2018). Physical activity classification with dynamic discriminative methods. *Biometrics* **74** 1502–1511. 2

REINER, M., NIERMANN, C., JEKAUC, D. and WOLL, A. (2013). Long-term health benefits of physical activity– a systematic review of longitudinal studies. *BMC public health* **13** 1–9. 1

ROBERTS, G. O. and ROSENTHAL, J. S. (2009). Examples of Adaptive MCMC. *Journal of Computational and Graphical Statistics* **18** 349–367. 11

SIKKA, R. S., BAER, M., RAJA, A., STUART, M. and TOMPKINS, M. (2019). Analytics In Sports Medicine: Implications And Responsibilities That Accompany The Era Of Big Data. *JBJS* **101** 276–283. 2

STAUDENMAYER, J., HE, S., HICKEY, A., SASAKI, J. and FREEDSON, P. (2015). Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. *Journal of applied physiology* **119** 396–403. 4, 21

STEIN, M. L., CHI, Z. and WELTY, L. J. (2004). Approximating Likelihoods For Large Spatial Data Sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66** 275–296. 9

TARALDSEN, K., CHASTIN, S. F., RIPHAGEN, I. I., VEREIJKEN, B. and HELBOSTAD, J. L. (2012). Physical Activity Monitoring By Use Of Accelerometer-Based Body-Worn Sensors In Older Adults: A Systematic Literature Review Of Current Knowledge And Applications. *Maturitas* **71** 13–19. 4

VAN HEES, V. T., RENSTRÖM, F., WRIGHT, A., GRADMARK, A., CATT, M., CHEN, K. Y., LÖF, M., BLUCK, L., POMEROY, J., WAREHAM, N. J. et al. (2011). Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. *PLoS one* **6** e22922. 4

VAN LOO, C. M., OKELY, A. D., BATTERHAM, M. J., HINKLEY, T., EKELUND, U., BRAGE, S., REILLY, J. J., TROST, S. G., JONES, R. A., JANSSEN, X. et al. (2018). Wrist acceleration cut-points for moderate-to-vigorous physical activity in youth. *Medicine and science in sports and exercise* **50** 609. 4

VECCHIA, A. V. (1988). Estimation And Model Identification For Continuous Spatial Processes. *Journal of the Royal Statistical Society: Series B (Methodological)* **50** 297–312. 9

WHITE, T., WESTGATE, K., WAREHAM, N. J. and BRAGE, S. (2016). Estimation of physical activity energy expenditure during free-living from wrist accelerometry in UK adults. *PLoS one* **11** e0167472. 4