Outlier detection in dynamic functional models

Andrada E. Ivanescu [1], William Checkley [2], and Ciprian M. Crainiceanu [3]

[1] School of Computing, Montclair State University, Montclair, NJ, USA

[2] Department of Medicine, Johns Hopkins University, Baltimore, MD, USA

[3] Department of Biostatistics, Johns Hopkins University, Baltimore, MD, USA

Address for correspondence: Ciprian M. Crainiceanu, Department of Biostatistics, Johns Hopkins University, 615 N. Wolfe Street, E3636, Baltimore, MD, 21205, USA.

E-mail: ccraini1@jhu.edu.

Phone: (+1) 410 955 3505.

Abstract: We propose methods for identification of dynamic outliers for sparse or dense functional data using dynamic z-scores. A dynamic outlier is defined as an unusual observation in the "future" of a trajectory given the "past" of that trajectory as well as the past and future of all other trajectories. Dynamic outlier identification methods can be used prospectively at any time point to estimate if and when the functional trajectory deviates from its predicted path. In contrast, static outliers are obtained after the entire data set is collected and can only be used retrospectively. We will illustrate the difference between dynamic outliers and static outliers using

extensive simulations studies and a longitudinal study of child growth conducted in Lima, Peru. In this study, dynamic outliers correspond to unusual patterns of growth for a child at any point of the study (unexpected changes), while static outliers correspond to unusually large or small children throughout the study.

Key words: child growth; cross-validation; dynamic outlier; dynamic regression; function-on-function regression; outliers.

1 Introduction

We propose new methods for dynamic outlier detection in the context of functional sparse or dense data using dynamic z-scores. A dynamic outlier is defined as an unusual observation in the "future" of a trajectory given the "past" of that trajectory as well as the past and future of all other trajectories. This is an important statistical problem motivated by a large number of scientific applications. For example, in human growth analysis one is interested in whether the latest height measurement for a child is unusual given the known history of growth of that child and of other children in the study. Detecting abnormal observations or patterns of growth could be used for early targeted health interventions and data quality control assessment. In many growth studies, in addition to the primary measurement (e.g., length of the baby) one often collects other longitudinal growth information (e.g., weight, head circumference) and baseline covariates (e.g., mother's social economic status at birth.) Thus, we define the "known history of the growth of a child" as the collection of all measurements up to a particular time point. From a technical perspective, this requires modeling a functional outcome as a function of its past (e.g., height of a child

measured in the first 9 months from birth), the past of other functional predictors (e.g., weight of the same child measured in the first 9 months from birth), the past and future of the trajectories of other study participants (e.g., the past and future growth measurements of other children), and additional covariates. The main differences between dynamic prediction and traditional time series forecasting are: (1) lack of stationarity assumptions; and (2) the dependence of the prediction on the past and future of the trajectories of the other study participants.

Given observations, $Y_{i,t}$, for study participant i and time t and any method that produces a predictor $\widehat{Y}_{i,t+h}$ at time $t+h$, $h>0$, and a standard deviation of prediction $\text{sd}(Y_{i,t+h}-\widehat{Y}_{i,t+h})$, the dynamic outliers score computed at time $t+h$ based on information up to time t is

$$\text{Z}_{i,t,h}=\frac{Y_{i,t+h}-\widehat{Y}_{i,t+h}}{\text{sd}(Y_{i,t+h}-\widehat{Y}_{i,t+h})}\ .$$

Our proposal is to calculate these z-scores for all study participants, $i=1,\ldots,I$, all time points t, and all prediction horizons, h, and identify the z-scores that exceed a certain threshold, say 3, 4, and 5. We propose to estimate both $\widehat{Y}_{i,t+h}$ and $\text{sd}(Y_{i,t+h}-\widehat{Y}_{i,t+h})$ using the dynamic functional regression models (Ivanescu et al. 2017), though other approaches could be used, as well. The idea is conceptually different, targets different types of outlying information, and leads to different results from existing, static, outlier detection methods. In particular, the approach is dynamic because it screens outliers at every observed time point compared to static approaches that use all available data. Static approaches are useful in retrospective studies when one is interested primarily in data quality control. In contrast, dynamic approaches are useful when one is interested in identifying unusual observations as soon as possible and use these findings for interventions *as data are acquired*.

Several methods for outlier detection in functional data have been proposed and use functional principal components (Sawant et al. 2012, Ren et al. 2017, Mejia et al. 2017), multivariate distances (Hyndman and Shang 2010, Hubert et al. 2015), and functional depth (Arribas-Gil and Romo 2014, Febrero et al. 2015). All these methods are static because they require the entire dataset to be collected, cannot include additional scalar or functional covariates, and have not been applied in a dynamic context.

Unusual growth patterns have been studied in the context of regression using functional principal component scores for univariate functional data (Zhang and Wei 2015), but methods have not been generalized to multivariate functional datasets (e.g., simultaneous growth in length, weight, and head circumference) or to the dynamic context, when one is interested in future unusual observations.

Our outlier detection approach is based on dynamic regression, where a new model is fit at every time point. Several methods (Goldberg et al. 2014, Chiou et al. 2016, Shang 2017a, Shang 2017b, Hyndman and Shang 2025) proposed inferential procedures for prediction using dynamic functional regression. In addition to proposing predictive models, some methods (Shang 2017a, Shang 2017b, Hyndman and Shang 2025) provided bootstrap-based confidence intervals for the predicted function; their method is implemented in the `ftsa` (Hyndman and Shang 2025) R package. However, these methods do not account for additional fixed effects or functional predictors.

The methodology we propose addresses a novel problem in functional data analysis where dynamic outliers can be detected. Methods start with using a class of dynamic prediction models to approximate the expected trajectory for an individual. Then, a novel metric, the dynamic z-score, is introduced and designed to detect departures

from the expected path. We ran novel simulation settings to test our proposed methods where data consists of functional datasets. We also performed new data analysis strategies for child growth studies. In applications and simulations we conducted, the numerical results for the proposed methods show that dynamic outliers were detected in functional or longitudinal data settings.

The paper offers important contributions to the functional data analysis research arena. The main strengths of our paper are: (1) conceptualizing and implementing a reasonable approach to detecting outliers dynamically that is novel and responds to a large and increasing number of scientific problems; (2) methods can be incorporated in practical research settings; (3) the methods and the problem are fundamentally different from the large existing literature on outlier detection in functional data; (4) the methods are practical, reasonable, and tested; and (5) we provide evidence that the methods detect outlying dynamics in real applications to child growth, which can have substantial impacts in analyses conducted in real time.

2 Methods

We consider the case when the data structure is of the type $Y_{i,l}, Z_{i,l}, X_i$ for $i = 1, \ldots, I$ and $l = 0, \ldots, t$. In our application $Y_{i,l}$ refers to the child's HAZ (height-for-age z-score) data observed at month 0 through t, $Z_{i,l}$ represents the WAZ (weight-for-age z-score) data for the same months, and X_i is the sex of the child. For simplicity we focus on the case when observations are sampled at the same time points across subjects, but methods apply to sparse, uneven sampling across subjects, as well. We are interested in two related problems. Given a sample of functions, $Y_{i,l}$, $i = 1, \ldots, I$, $l = 0, \ldots, t$ and for every time point $t + h \in \{t + 1, \ldots, 15\}$ we would like to identify

and quantify: (1) outlying observations $Y_{i,t+h}$ at time $t+h > t$; and (2) outlying patterns $\{Y_{i,t+h} : t+h > t\}$.

2.1 Dynamic prediction

Denote by $\widehat{Y}_{i,t+h} = E(Y_{i,t+h}|X_i, Y_{i,l}, Z_{i,l})$ the predicted "future" trajectory for subject i, which depends on the past of the time-varying processes $Y(\cdot)$ and $Z(\cdot)$ as well as on the time-invariant covariate X_i. We are interested in conditional models of the type $\widehat{Y}_{i,t+h} = f(X_i, Y_{i,l}, Z_{i,l})$, where $f(\cdot)$ denotes a function whose form will be specified and will be modeled either parametrically or semiparametrically. There are many options for building predictive models via specifications of the function $f(\cdot)$. Here we use the strategy (Ivanescu et al. 2017) that considers a class of explicit regression models. To ensure that the presentation is self-contained we re-describe this class here. The simplest model considered is the BENchmark DYnamic (BENDY) model

$$Y_{i,t+h} = X_i\gamma_{t+h,t} + Y_{i,0}\beta_{0,t+h,t} + Y_{i,t}\beta_{t,t+h,t} + Z_{i,0}\delta_{0,t+h,t} + Z_{i,t}\delta_{t,t+h,t} + \epsilon_{i,t,h} \ . \quad (2.1)$$

The model parameters depend on when the prediction is performed, t, and at what time into the future, $t+h$. This explicit dependence ensures that predictions are targeted to the time point where predictions are produced and use the available historical data information. A potential criticism of model (2.1) could be that the outcome may not depend only on the last and first available data points, but also on the observations between them. To address this, we extend (2.1) to the class of Dynamic Linear Models (DLM)

$$Y_{i,t+h} = X_i\gamma_{t+h,t} + \sum_{l=0}^{t} Y_{i,l}\beta_{l,t+h,t} + \sum_{l=0}^{t} Z_{i,l}\delta_{l,t+h,t} + \epsilon_{i,t,h} \ . \quad (2.2)$$

The model has a functional data interpretation as well, and can be written as

$$Y_{i,t+h} = X_i\gamma_{t+h,t} + \int_{l=0}^{t} Y_{i,l}\beta_{l,t+h,t}dl + \int_{l=0}^{t} Z_{i,l}\delta_{l,t+h,t}dl + \epsilon_{i,t,h} \ . \quad (2.3)$$

In practice, model (2.2) can be unstable due to the high correlation among predictors. To alleviate this problem one can impose quadratic penalties on the model parameters $\beta_{l,t+h,t}$ and $\delta_{l,t+h,t}$. More precisely, DPFR imposes a different quadratic penalty on $\beta_{l,t+h,t}$ for every fixed t and h, whereas DPFFR imposes a quadratic penalty for every fixed t. Separate prediction models are fit for each (t, h) pair for BENDY, DLM, and DPFR. The only exception is DPFFR where one model is fit at month t.

2.2 Dynamic z-scores

In this section we describe the modelling strategy for constructing dynamic z-scores. The term that involves the historic data $Y_{i,l}$ from equation (2.3) contains the model parameter $\beta_{l,t+h,t}$, which is assumed to be a smooth function over l and that depends on the historic time point l and lag time h. All our models assume that this function can be expanded as $\beta_{l,t+h,t} = \sum_{k=1}^{K_1} B_{1,k}(l, t+h)a_{k,t}$, where $B_{1,k}(\cdot,\cdot)$ is a bivariate basis evaluated at points l (past of the trajectory) in the first dimension, and times $t+h$ (future of the trajectory) in the second. Thus, the integral $\int_{l=0}^{t} Y_{i,l}\beta_{l,t+h,t}dl$ in model (2.3) can be approximated by the Riemann sum approximation

$$\sum_{k=1}^{K_1}\left\{\sum_{l=0}^{t} B_{1,k}(l,t+h)\widetilde{Y}_{i,l}\right\}a_{k,t} = \sum_{k=1}^{K_1} A_{1,i,k,t+h}a_{k,t} \ ,$$

where $\widetilde{Y}_{i,l} = \Delta_l Y_{i,l}$ and Δ_l is the length of the intervals in the Riemann approximation. The quantities $A_{1,i,k,t+h} = \sum_{l=0}^{t} B_{1,k}(l,t+h)\widetilde{Y}_{i,l}$ are known because $Y_{i,l}$ are considered observed for $l \leq t$, and $B_{1,k}(\cdot,\cdot)$ are known basis functions. Similarly, the term $\int_{l=0}^{t} Z_{i,l}\delta_{l,t+h,t}dl$ from equation (2.3) can be approximated by the Riemann sum

$$\sum_{k=1}^{K_2}\left\{\sum_{l=0}^{t} B_{2,k}(l,t+h)\tilde{Z}_{i,l}\right\}b_{k,t} = \sum_{k=1}^{K_2} A_{2,i,k,t+h}b_{k,t} \ .$$

Therefore, model (2.3) can be approximated by the following additive model

$$Y_{i,t+h} = X_i\gamma_{t+h,t} + \sum_{k=1}^{K_1} A_{1,i,k,t+h}a_{k,t} + \sum_{k=1}^{K_2} A_{2,i,k,t+h}b_{k,t} + \epsilon_{i,t,h} \ . \tag{2.4}$$

This model is estimated using penalized spline estimation based on the criterion

$$\sum_{i,h} ||Y_{i,t+h} - \mu_i(t+h;\boldsymbol{\gamma},\boldsymbol{a},\boldsymbol{b})||^2 + \lambda_\beta P_a(\boldsymbol{a}) + \lambda_\delta P_b(\boldsymbol{b}) \ , \tag{2.5}$$

where $\mu_i(t+h;\boldsymbol{\gamma},\boldsymbol{a},\boldsymbol{b})$ is the mean of $Y_{i,t+h}$ and $\boldsymbol{\gamma},\boldsymbol{a},\boldsymbol{b}$ are the vectors containing the corresponding parameters. The penalized criterion in equation (2.5) can be shown to be equivalent to a specific mixed effects model, which allows to conduct inference using standardized software; see Crainiceanu et al. 2024, Ch. 6.2.1 for more details.

Irrespective of the model fit, $Y_{i,t+h} = \widehat{Y}_{i,t+h} + \widehat{\epsilon}_{i,t+h}$ and, under the normality assumption of parameter estimators in model (2.3), $\widehat{Y}_{i,t+h}$ is also normal at each time point $t+h$ because these expressions are conditional on the past observations, $Y_{i,l}$ and $Z_{i,l}$, $l = 0,\ldots,t$. We define the z-scores $\mathrm{Z}_{i,t,h} = (Y_{i,t+h} - \widehat{Y}_{i,t+h})/\sqrt{\mathrm{Var}\{\widehat{Y}_{i,t+h}\} + \mathrm{Var}\{\epsilon_{i,t,h}\}}$ where $Y_{i,t+h}$ is the observed data for child i at time $t+h$, $\widehat{Y}_{i,t+h}$ is the prediction of $Y_{i,t+h}$ for child i given their data history.

The predictions $\widehat{Y}_{i,t+h}$ are obtained based on the corresponding models using published methods and code. For BENDY and DLM calculations are based on the standard prediction procedures for parametric regression. For DPFR and DPFFR calculations are based on best linear unbiased prediction (BLUP) in the associated mixed effects representation of penalized regression (Crainiceanu et al. 2024, Goldsmith et al. 2011, Scheipl et al. 2015). Our previous work allowed us to implement these methods using stable software such as the `refund` (Goldsmith et al. 2025) and `mgcv` (Wood 2025) packages in `R` (R Core Team 2025). The dynamic z-score can be compared to the Normal or t-quantiles, though here we use the more conservative choice $q_{16-(t+1),\alpha}$,

which is the $1-\alpha/2$ quantile of the t distribution with $16-(t+1)$ degrees of freedom. We used $t+1$ degrees of freedom because the length of the history used for variance estimation is relatively short. Here $\text{Var}(\widehat{Y}_{i,t+h})$ quantifies the variability of the prediction of the true trajectory, whereas $\text{Var}(\epsilon_{i,t,h})$ quantifies the additional, irreducible, variability associated with observations around the long term trend. Note that one can also produce the symmetric $100(1-\alpha)\%$ point-wise confidence interval for future observations $Y_{i,t+h}$ as $\widehat{Y}_{i,t+h} \pm q_{16-(t+1),\alpha}\sqrt{\text{Var}\{\widehat{Y}_{i,t+h}\} + \text{Var}\{\epsilon_{i,t,h}\}}$.

Obtaining the variance estimates from these models requires some additional effort, but all necessary quantities are readily available. More precisely, for either DPFR or DPFFR we will use the sandwich variance estimator $\widehat{\text{Var}}(\widehat{Y}_{i,t+h}) = P_i \widehat{V} P_i^t$, where $P_i = [X_i, \{Y_{i,l}\}_{l=0}^t, \{Z_{i,l}\}_{l=0}^t]$ is a row vector that consists of the predictive information for child i up to time t and $\widehat{V}$ is the estimated covariance of the model parameters. We use Bayesian shrinkage estimation of the fixed and random effect model parameters (Wood 2025, Wood 2006), while the variance of the residuals is estimated using the average of the squared estimated residuals.

3 Application to the CONTENT study

The CONTENT study is a child growth study conducted in Pampas de San Juan Miraflores and Nuevo Paraiso, two peri-urban shanty towns with high population density, 25 km south of central Lima, Peru. These peri-urban communities (Checkley et al. 1998, Checkley et al. 2002) are comprised of 50,000 residents, the majority of whom are immigrants from rural areas of the Peruvian Andes who settled nearly 35 years ago and later claimed unused land on the outskirts of Lima. In the last two decades, Pampas has undergone many economic and social developments. In 1989, most homes

were temporary structures, constructed of wooden poles and woven thatch, without water or sewage lines. Currently, more than 85% of homes are constructed from brick or cement with in-home water and sewage lines. The study was approved by the European Union Ethics Committee, A.B. PRISMA and Universidad Peruana Cayetano Heredia, in Lima, Peru, and the Bloomberg School of Public Health, Johns Hopkins University, in Baltimore, USA. Mothers or caregivers provided written informed consent. Dynamic functional regression models were applied to the CONTENT data to identify potential dynamic outliers. There were a total of $I = 197$ babies in the dataset (Grajeda et al. 2016, Ivanescu et al. 2017, Leroux et al. 2018, Crainiceanu et al. 2024). Dynamic identification of outliers in the HAZ (height-for-age z-score) data is the primary focus of our analysis. Given the HAZ history of a child, dynamic outlier identification for HAZ was conducted using four dynamic prediction (Ivanescu et al. 2017) methods: BENDY, DLM, DPFR, and DPFFR. The CONTENT dataset was made available in `refund` (Goldsmith et al. 2025) and several analyses are presented on the website www.functionaldataanalysis.org accompanying the recent monograph on FDA with R (Crainiceanu et al. 2024).

3.1 Data imputation

The CONTENT data set contains measurements made at unequal intervals, with more observations immediately after birth and fewer as time progressed. A first step was the computation of the monthly average value for HAZ and WAZ. A second step was to reconstruct the individual trajectories and impute observations for the months with no observations. This was done using functional principal components method (Goldsmith et al. 2013), which both smoothed the data trajectories and provided 16 observations for each child, one observation per month. Noise was added back to

the reconstructed FPCA trajectories by: (1) computing the variance of the observed residuals at each month separately and taking the minimum variance across months; and (2) simulating independent random errors from a zero-mean normal with this variance and adding them to the smooth trajectories. Figure 1 displays the side by side comparison of the data before and after imputation for both HAZ and WAZ. The resulting data set is highly realistic and mimics the main patterns in the original data. All methods presented in this paper refer to the complete HAZ and WAZ data shown in the second column of Figure 1.

3.2 Covariate unadjusted outlier identification

In this section several case studies are discussed based on the analysis of the HAZ data without additional covariates, while Section 3.3 provides results for covariate-adjusted analyses. The section begins with identifying subjects with outlying HAZ dynamic z-score values. For simplicity we present results based on historic data from birth until month 3, and from birth until month 7, respectively. However, methods can and are applied to every point in the history of the growth curves.

When using the first three months as historical data for prediction of the future trajectory we obtained 8% outliers when using a threshold for dynamic z-scores of -4 and 15% outliers when using a threshold of -3. When using the first seven months as historical data we obtained 0.12% outliers for a z-score threshold of -4 and 1.2% outliers when using a threshold of -3. The percent number of outliers was obtained by dividing the total number of outliers by the number of observations for all children at all future points.

For illustration Table 1 displays all outliers for two children (IDs 86 and 135) by

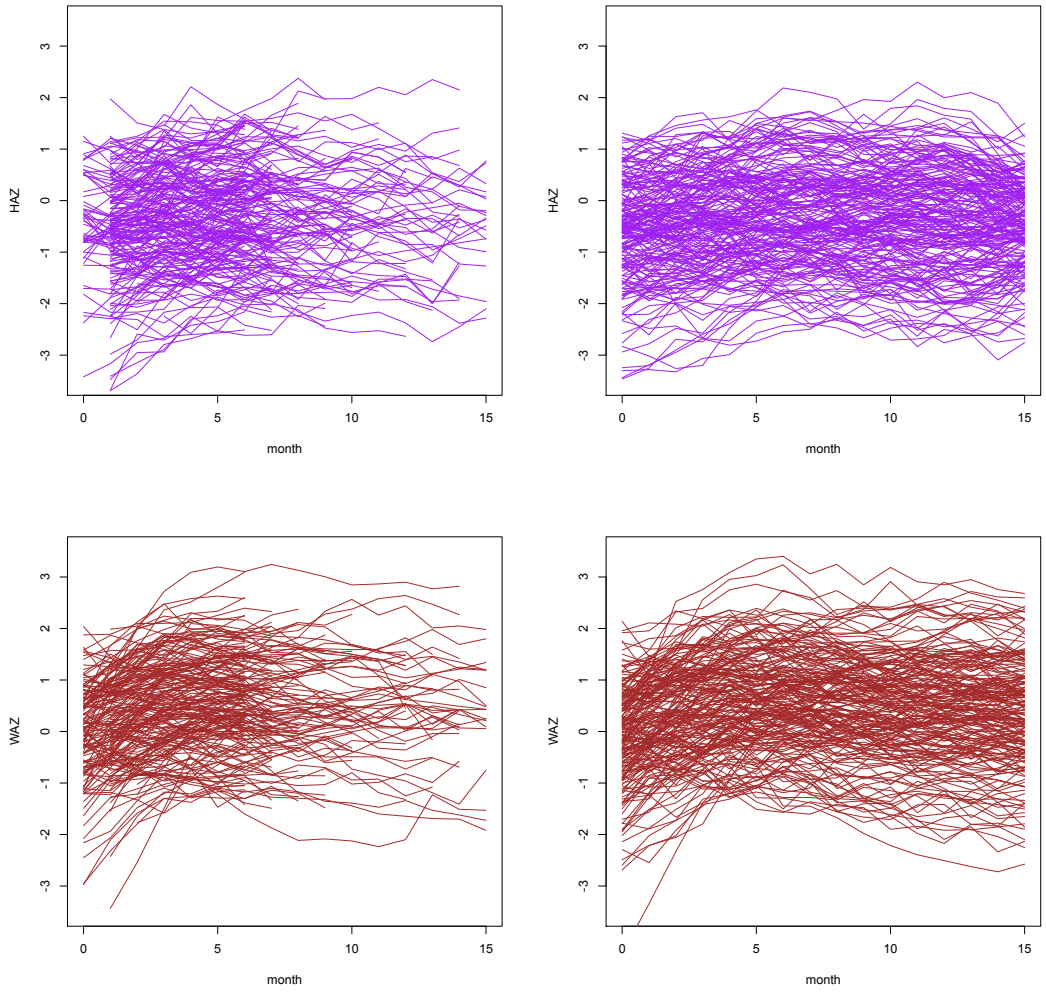

Figure 1: HAZ (first row) and WAZ (second row) for the original data (first column) and reconstructed data (second column).

method, time at the time of prediction, and time in the future where prediction is conducted. Outliers were detected using the quantile $q_{16-(t+1),\alpha}$ discussed in Section 2.2. Results indicate that methods tend to be in agreement irrespective of the model used for data fitting. To better understand these results, consider the child with data shown in the first row in Figure 2 (ID 86). This is a child who starts close to the World Health Organization (WHO) average of HAZ (-0.6 HAZ at baseline)

Table 1: Dynamic identification of HAZ outliers for 2 children. Numbers in blue represent the months between 8-15 with HAZ outliers identified at times 3 and 7. Numbers in black represent the corresponding dynamic z-scores.

Covariate-unadjusted	$t = 3$	$t = 7$
ID 86		
DPFFR	8, 9, 10, 11, 12, 13, 14, 15 −6.6 −6.7 −6.8 −6.7 −6.5 −5.7 −4.8 −3.6	8, 9, 10, 12 −2.3 −2.4 −2.4 −2.3
DPFR	8, 9, 10, 11, 12, 13, 14, 15 −6.2 −5.8 −6.0 −6.2 −6.1 −5.7 −4.3 −3.2	8 −2.4
DLM	8, 9, 10, 11, 12, 13, 14, 15 −6.2 −5.8 −6.0 −6.2 −6.1 −5.7 −4.4 −3.2	8, 12, 13 −2.5 −2.3 −2.4
BENDY	8, 9, 10, 11, 12, 13, 14, 15 −5.7 −5.6 −5.9 −5.8 −5.8 −5.2 −4.3 −3.3	8, 9, 10, 11, 12, 13, 14, 15 −2.4 −2.6 −2.9 −2.9 −3.1 −2.9 −2.7 −2.6
ID 135		
DPFFR	10, 12, 13, 14 −2.6 −2.6 −3.5 −3.3	10, 12, 13, 14 −2.8 −2.8 −3.7 −3.4
DPFR	12, 13, 14 −2.4 −3.3 −3.0	10, 12, 13, 14 −2.4 −2.8 −3.7 −3.2
DLM	12, 13, 14 −2.3 −3.3 −2.9	10, 12, 13, 14 −2.5 −2.8 −3.6 −3.2
BENDY	13, 14 −3.0 −2.8	13 −2.4

and stays there for the first several months. However, after month 7 the HAZ (static z-score) decreases until about month 13, even though it remains above -1.5, and would not trigger an intervention. However, all dynamic models indicate that the dynamic z-scores for $t = 3$ and $t+h = 8, \ldots, 15$ are negative and very large in absolute value, indicating that investigation and, possibly, intervention may be necessary. This indicates that, at least for this child, the method provides a credible early warning of stunting behavior. Interestingly, the observations at months 8 through 10, and 12, are identified as outlying using both historical data up to month 3 and month 7. However, using data only up to month 3 identifies a few additional outliers (11, 13, 14 and 15). The agreement is reassuring, while the difference could be explained as follows. Having the four initial observations of HAZ (-0.6, -0.8, -0.5, and -0.2) is inconsistent with HAZ values of -1.2, -1.1, and -1.0 at months 13, 14 and 15,

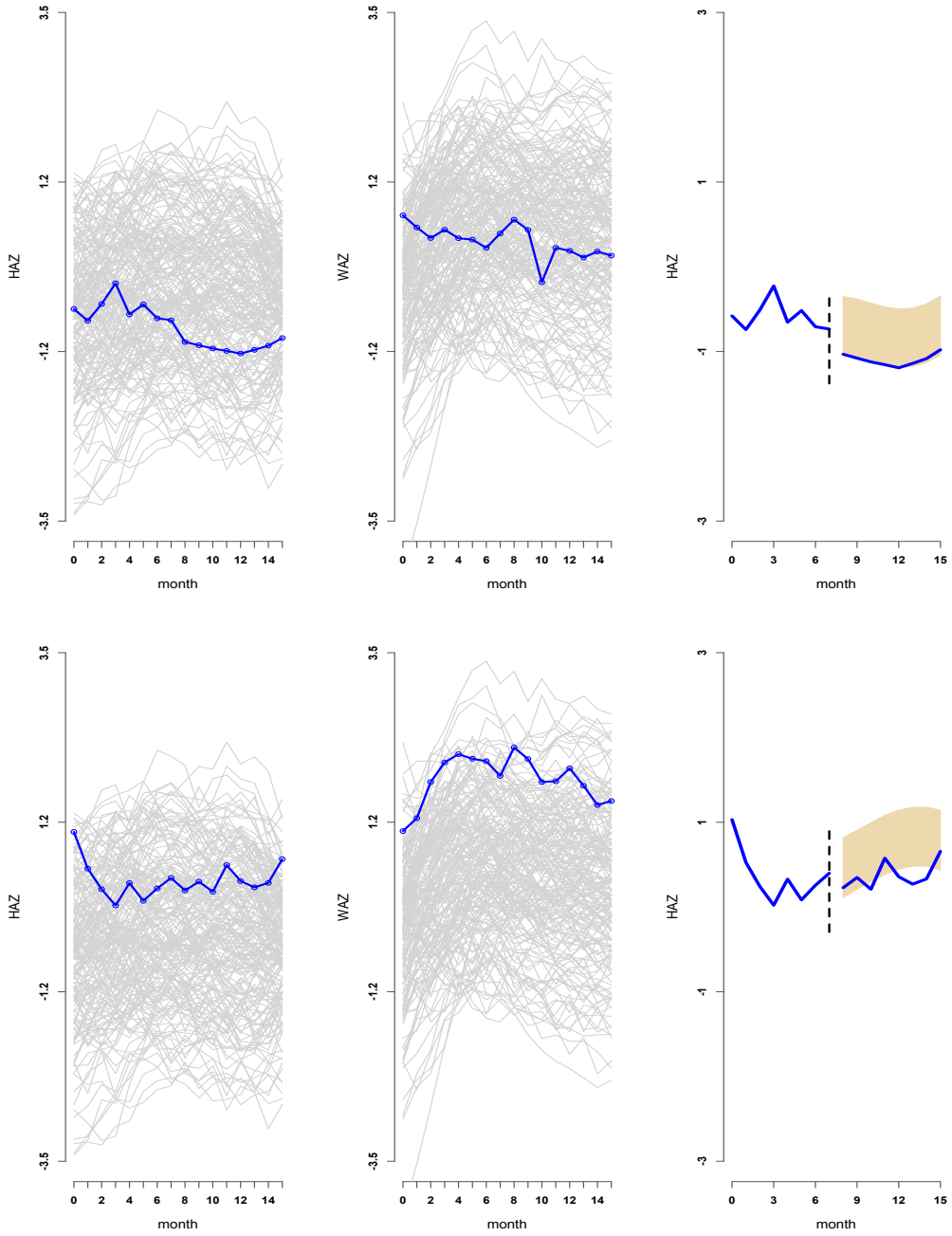

Figure 2: Application to child growth data. HAZ trajectories (left column), WAZ trajectories (middle column) and HAZ dynamic prediction intervals in orange (right column) for $t = 7$ corresponding to study participant ID 86 (first row) and ID 135 (second row). Observed data shown in blue solid line. Scenario: Covariate-adjusted.

respectively, whereas these values are not considered that extreme given the additional observations in the history up to month 7 (-0.7, -0.5, -0.7, and -0.7). We contend that using dynamic z-scores provides quantification to a very natural question in this type of studies: given the data for an individual up to time t and data for everyone else up to time $t+h$, what new observations are outlying. This is different from time series forecasting, because data information about the future is borrowed from the trajectory of other individuals and stationarity is not required.

3.3 Covariate adjusted outlier identification

The results of the previous section were based solely on the HAZ history. We now focus on identifying HAZ outliers based on HAZ and WAZ history as well as sex. The effect of the sex covariate was modeled as linear effect of time and BENDY, DLM, DPFR and DPFFR were used to calculate dynamic z-scores. Table 2 displays the same results as Table 1, but for the case when we further adjust for the dynamic history of WAZ as well as sex.

The adjusted and unadjusted analyses provide qualitatively the same results. For example, for the first child (ID 86) the months in the future identified as outliers are almost identical based on data at 3 months (compare Tables 1 and 2). For the second child (ID 135) the months when outliers are identified agree closely when predicting at month $t = 7$, with more differences when predicting at time $t = 3$. In this case, all covariate-adjusted models, except DPFR, identify more outliers than the unadjusted analyses. This could be due to the very short history ($t = 3$, hence 4 data points) and possible instabilities of the fitting methods for such extreme cases.

The right panels in Figure 2 illustrates the prediction results based on the DPFFR

Table 2: Dynamic identification of HAZ outliers for 2 children (covariate-adjusted). Numbers in blue represent the months between 8-15 with HAZ outliers identified at times 3 and 7. Numbers in black represent the corresponding dynamic z-scores

Covariate-adjusted	$t = 3$	$t = 7$
ID 86		
DPFFR	8, 9, 10, 11, 12, 13, 14, 15 −4.9 −5.0 −5.1 −5.1 −5.1 −4.5 −3.9 −3.0	9, 10, 11, 12 −2.4 −2.3 −2.3 −2.4
DPFR	8, 9, 10, 11, 12, 13, 14, 15 −4.9 −5.4 −5.7 −5.2 −5.4 −5.6 −4.2 −3.6	8, 10, 11, 12, 13 −3.3 −2.5 −2.7 −2.5 −2.4
DLM	8, 9, 10, 11, 12, 13, 14, 15 −4.8 −4.3 −4.4 −4.5 −4.5 −4.2 −3.2 −2.7	8, 13 −2.4 −2.4
BENDY	8, 9, 10, 11, 12, 13, 14, 15 −4.0 −3.8 −4.1 −3.8 −3.8 −3.4 −2.9 −2.7	8, 9, 10, 11, 12, 13, 15 −2.4 −2.4 −2.7 −2.5 −2.7 −2.4 −2.3
ID 135		
DPFFR	8, 9, 10, 11, 12, 13, 14 −2.5 −2.4 −4.0 −2.2 −4.0 −4.7 −4.3	10, 12, 13, 14 −2.9 −2.9 −3.7 −3.3
DPFR		8, 10, 13 −3.5 −2.8 −2.5
DLM	8, 10, 12, 13, 14 −2.2 −2.8 −3.6 −4.3 −3.6	12, 13, 14 −2.9 −3.7 −3.0
BENDY	10, 12, 13, 14 −2.3 −3.0 −3.7 −3.3	10, 13, −2.3 −2.9

covariate-adjusted method for the two children (IDs 86 and 135) based on data up to month $t = 7$. The orange areas are the corresponding 95% confidence intervals. The dynamic prediction intervals provide a helpful visual tool for a fast identification of dynamic outliers, though they depend on the choice of the confidence interval, whereas the dynamic z-scores are independent of the level of the confidence interval.

3.4 Comparisons with static outlier detection methods

There are many methods for identifying outliers in functional data and several of them have software implementations. These methods are not directly comparable to our approach because they are static, retrospective, and do not include additional time-invariant or time-dependent covariates. Here we will show that there are major differences between static and dynamic outliers methods in practice. Comparing the

methods is not straightforward, as static methods focus on identifying an entire curve and do not include covariates, whereas our methods are focused on identifying specific outliers given a past history and can include covariates. Regardless, we will apply the static methods and compare them with the new, dynamic, method introduced in this paper. More specifically, we fix a length of the history, say 7 months, and, for each subject, we calculate the average absolute value of the dynamic z-score for months 8 through 15. Study participants are then ranked in decreasing order from the largest to the smallest average absolute z-score. We repeat the procedure for the maximum absolute z-score and the number of outliers (absolute z-scores larger than the quantile $q_{16-(t+1),\alpha}$, though stricter definitions could be used, as well).

We compare our methods with static outlier detection approaches designed for multivariate functional data. In particular, we focus on methods based on the functional band depth for univariate functional data (Lopez-Pintado and Romo 2009) and extended to multivariate functional data (Ieva and Paganoni 2013, Ieva et al. 2019). The corresponding functions in `R` are `multiMBD` and `multiBD`, where "multi" stands for multivariate functional data, "BD" stands for band depth and "MBD" stands for modified band depth. These functions are implemented in the `roahd` (Ieva et al. 2019) R package. We also compare our approach to the outliergram (Arribas-Gil and Romo 2014) approach introduced for univariate functional data. The corresponding function is `outliergram` implemented in the `roahd` (Ieva et al. 2019) R package that also provides extensions to multivariate functional data. The `roahd` (Ieva et al. 2019) R package offers an implementation version of `multivariate_outliergram` for multivariate functional data. For the multivariate outliergram approach (referred as `moutliergram` in our study) we used the default critical F-value of 1.5, which resulted in 14 outliers, while an F-value of 2, identified 10 outliers. The band depth

approaches do not offer a cut-off point for defining outliers. Therefore, we ranked the depths for every subject and retained a number of subjects equal to the number of subjects identified by the multivariate outliergram approach.

For our DPFFR dynamic outlier detection we distinguish three approaches for identifying subjects with large dynamic outliers: $\text{DPFFR}_{\text{ave}}$, $\text{DPFFR}_{\text{max}}$, and DPFFR_{nr}, corresponding to the average, maximum, and number of dynamic z-scores with an absolute value greater than the t quantile $q_{16-(t+1),\alpha}$, respectively. For illustration purposes we use historic data up to $t = 7$ months and relied on the DPFFR dynamic z-scores calculated at all future months $t + h \in \{t + 1, \ldots, 15\}$. The top 10 outlying individuals for each method were compared and the confusion matrix is provided in Table 3. There were 37 different IDs identified in the top 10 across all methods presented. For example, there were 18 individual curves that were not identified as outliers by `moutliergram` or $\text{DPFFR}_{\text{max}}$ (indicated as $(0, 0)$), there were 9 outliers identified by `moutliergram` but not by $\text{DPFFR}_{\text{max}}$ (indicated as $(1, 0)$), there were 9 outliers identified by $\text{DPFFR}_{\text{max}}$ but not by `moutliergram` (indicated as $(0, 1)$), and there was only 1 outlier identified by both methods.

Table 3 indicates that: (1) static and dynamic methods do not agree in terms of identifying outliers; (2) $\text{DPFFR}_{\text{ave}}$ and $\text{DPFFR}_{\text{max}}$, as well as $\text{DPFFR}_{\text{ave}}$ and DPFFR_{nr}, have the highest agreement among DPFFR-derived methods for outlier detection; and (3) some static methods agree quite well among themselves. We emphasize that this is not how we propose to use our outlier detection algorithms; this was done only for comparing to existing static methods, which required substantial changes to our procedure to even obtain something that is comparable. To summarize, dynamic and static outlier identification are different both in terms of how they are defined and

Table 3: Results for outlier detection by several static methods and dynamic detection at $t = 7$ with DPFFR. Results "0" and "1" refer to the outlier detection results (1=detected, 0=not detected).

		Methods									
		multiMBD		multiBD		$\text{DPFFR}_{\text{ave}}$		$\text{DPFFR}_{\text{max}}$		DPFFR_{nr}	
		0	1	0	1	0	1	0	1	0	1
moutliergram											
	0	17	10	17	10	17	10	18	9	18	9
	1	10	**0**	10	**0**	10	**0**	9	**1**	9	**1**
multiMBD											
	0			23	4	18	9	18	9	17	10
	1			4	**6**	9	**1**	9	**1**	10	**0**
multiBD											
	0					18	9	18	9	17	10
	1					9	**1**	9	**1**	10	**0**
$\text{DPFFR}_{\text{ave}}$											
	0							24	3	24	3
	1							3	**7**	3	**7**
$\text{DPFFR}_{\text{max}}$											
	0									22	5
	1									5	**5**

used and in terms of their results. More information about the top 10 outliers are posted in the appendix material.

To better understand the differences between the type of outliers identified by the different methods, we now visualize some of these outliers. The child with ID 17 was identified as an outlier by static and dynamic methods, but for different reasons. Indeed, the HAZ and WAZ trajectories for this study participant are shown as blue solid lines in Figure 3. This study participant had high values for WAZ for almost the entire time range, except for the first two months. The HAZ values go up for the first 6 months, as well, from 0.8 to about 1.5, but they go down again to below 0.8 at month 15. Dynamic methods identify this subject as an outlier because the decrease in HAZ after month 7 is very unusual given the HAZ trajectory up to month

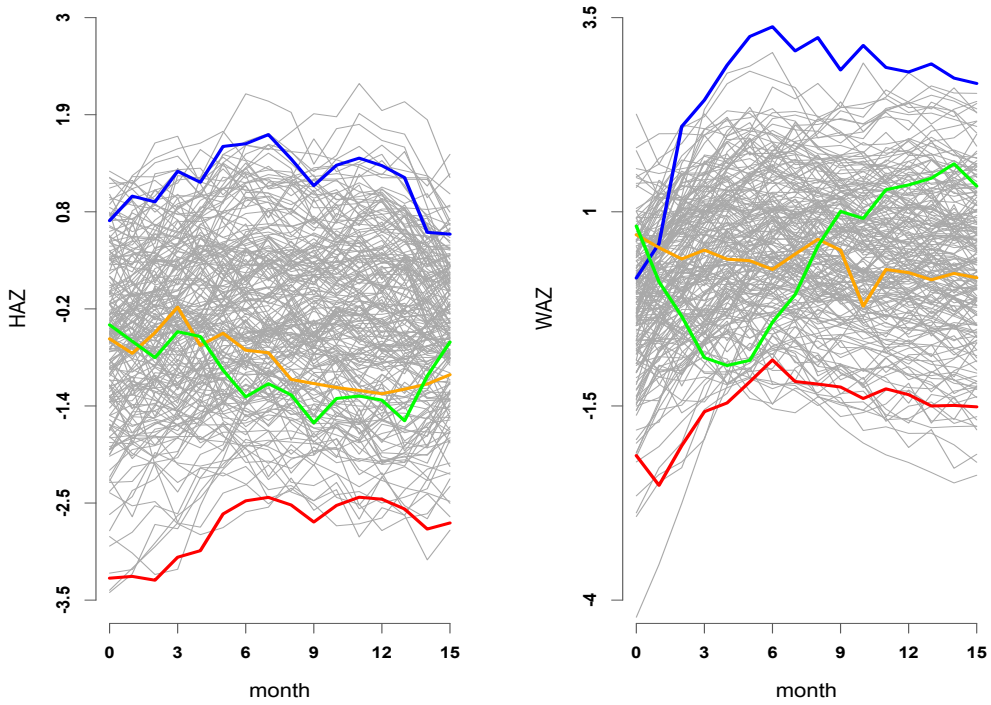

Figure 3: HAZ and WAZ data (gray lines) for 197 children. The highlighted curves correspond to data observed for four children (ID 17 in blue, ID 86 in orange, ID 32 in red, and ID 132 in green).

7 and the continuous increase in WAZ. Static methods pick up these unusual patterns because the child was the heaviest (at least after month 7) and among the tallest in this group. However, dynamic methods would detect the unusual pattern of growth at month 8 and 9, when information is actionable, whereas static methods would need to wait until month 15, when it is too late for an intervention. The second example is the child with ID 86, whose data are displayed as orange solid lines in Figure 3. This study participant was identified as a strong outlier by the dynamic approach, but was not identified as an outlier by any of the static methods (rank 164 for `multiMBD` and 142 for `multiBD` out of 197 children, where rank 1 is most and 197 is least outlying). This child's trajectory is characterized by a decreasing HAZ trend

after an initial increase in HAZ. This led to a predicted HAZ that was consistently higher than the actual observed HAZ values. This result indicates that static methods are measuring something fundamentally different from dynamic models, at least for this subject. The third example is the child with ID 32, whose data is shown as solid red lines in Figure 3. This study participant was identified as an outlier by two of the static methods (rank 1 for `multiMBD` and rank 4 by `multiBD`) but was not identified by any of the dynamic outlier detection methods. The reason is that this is a child who is among the lightest and shortest in the cohort, but their growth pattern is not identified as abnormal. The fourth example is the child with ID 132, whose data is shown in green solid lines in Figure 3. This study participant was identified by the multivariate outliergram method as a shape outlier. This may be due to the WAZ data that exhibited a sharp decline followed by a sharp rise in WAZ data in the first 9 months. For this child the HAZ data indicated a possible stunting pattern, as a dynamic z-score of -3.58 was observed at month 9.

4 Simulation study

In this section we focus on comparing the performance of the dynamic and static outlier identification methods when two different types of dynamic outliers of varying magnitude are added to realizations of Gaussian Processes. Data sets were initially simulated without outliers from the following data generating mechanism

$$Y_i(t) = \sum_{k=1}^{4} \xi_{i,k}\phi_k(t) + \epsilon_{i,t} \ . \tag{4.1}$$

We used the functions $\phi_1(t) = \sqrt{1.5}\sin(.875\pi t)$, $\phi_2(t) = \sqrt{1.5}\cos(.875\pi t)$, $\phi_3(t) = \sqrt{1.5}\sin(\pi t)$, and $\phi_4(t) = \sqrt{1.5}\cos(\pi t)$. The coefficients were simulated as independent $\xi_{i,k} \sim N(0, \lambda_k)$, where $\lambda_k = 0.5^{k-1}$ for $k = 1, 2, 3, 4$. Functions were simulated across

the equally spaced grid $t_m = \frac{m}{100}$ for $m = \{0, 1, \ldots, 100\}$ and each simulation included $I = 200$ curves. Different levels of noise were simulated, and we summarize results for the case $\epsilon_{i,t_m} \sim N(0, \sigma_\epsilon^2)$. We took $\sigma_\epsilon = 0.4$ for Section 4.1 and this value was lowered to $\sigma_\epsilon = 0.1$ for Section 4.2. In Sections 4.1 and 4.2 we provide simulation results for detecting two types of outliers, respectively. Within each section, the magnitude of each type of outlier is varied to cover a wide range, from very small to very large deviations. We contend that these scenarios are an important contribution in themselves as they provide the first simulations of this type.

4.1 Simulation of type I dynamic outliers

Once the data was generated from model (4.1), a trajectory i was sampled at random and a dynamic outlying trajectory was induced, as described below. More precisely, the constant $C \in \{-1.5, -1, -0.8, -0.4\}$ was added to $Y_i(t_m)$ for $t_m = \frac{m}{100}$ for $m \geq 59$. To visualize the size of the jump induced in the data, Figure 4 displays four simulated data sets (gray lines) with one outlier trajectory added (original trajectory shown in blue, modified trajectory shown in red). The panels are organized from the largest ($C = -1.5$ in the first row, first column) to the smallest jump ($C = -0.4$ second column, second row). The outlier starts at $t_m = 0.59$, but this choice is for illustration purposes only. The magnitude of the outlier depends on the range of the data, the size of the noise, and the complexity of signal. Note that if the two trajectories were not highlighted, one could probably not detect the unusual pattern shown in red by simply inspecting the gray plots. Instead, one would be much more focused on the extreme trajectories either at the top or bottom of the panels. Dynamic outlier identification methods introduced in this paper are designed to detect structural changes in individual trajectories, as shown in Figure 4. Existing, static

outlier detection models are not designed with the same aim. Our simulations will show that, indeed, dynamic and static outlier detection methods perform differently. This should not be surprising, as they are designed for different purposes and use different optimization criteria.

4.1.1 Results

For each magnitude of deviations $C \in \{-1.5, -1, -0.8, -0.4\}$ a total of $B = 300$ datasets were simulated, each containing $I = 200$ trajectories. For each data set, dynamic modeling was performed using BENDY, DLM, DPFR, and DPFFR, the dynamic z-scores were calculated for each method and time point $t_m \geq 0.59$, and then the absolute value of the z-scores were averaged within the method for $t_m \geq 0.59$. More sophisticated approaches could be used, but here we keep things simple.

Results are summarized in Figure 5 by method (each method is indicated by a different shade of red) and jump size, C, from the highest jump $C = -1.5$ (left) to the lowest jump $C = -0.4$ (right). Each boxplot displays the average absolute value of the z-score for all values of the function between 0.59 and 1; there are a total of 300 such observations for each boxplot corresponding to the number of simulations. Methods are all applied on the same data sets to make results comparable. As expected, across methods, as the size of the jump C decreases, the average absolute z-score decreases indicating more power to detect jumps. Indeed, for a fixed shade of red compare boxplots from left to right to notice this trend. Some methods, such as DPFR and DPFFR, perform consistently better, while DLM performs worse (compare the boxplots in darker versus lighter shades of red.) BENDY performs better than DLM in terms of the median, though it is also much more variable (compare the lightest shade of red to the next lightest shade of red).

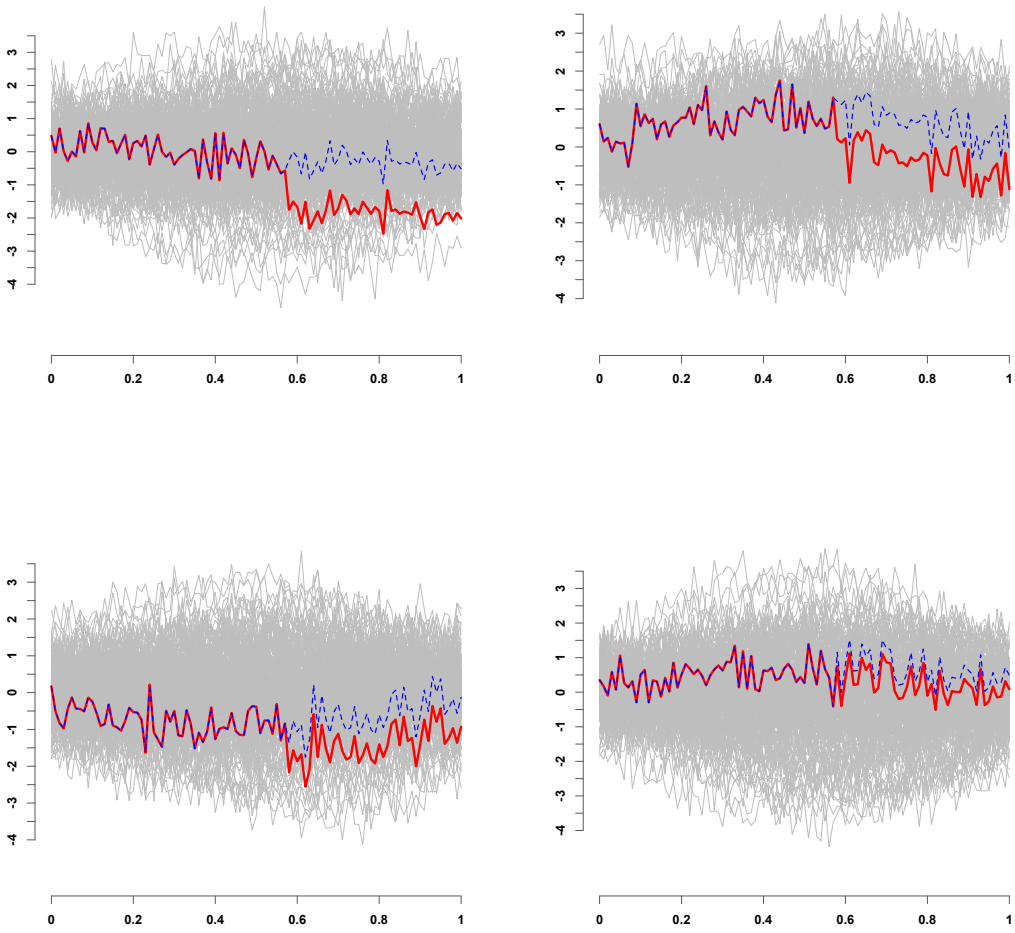

Figure 4: Four simulated functional datasets with $I = 200$ trajectories (light gray). The simulated dynamic outlier (red line) is obtained by adding a constant, C, to the original trajectory (dashed blue line). Scenario: $C = -1.5$ (first row, left panel), $C = -1$ (first row, right panel), $C = -0.8$ (second row, left panel), and $C = -0.4$ (second row, right panel)

.

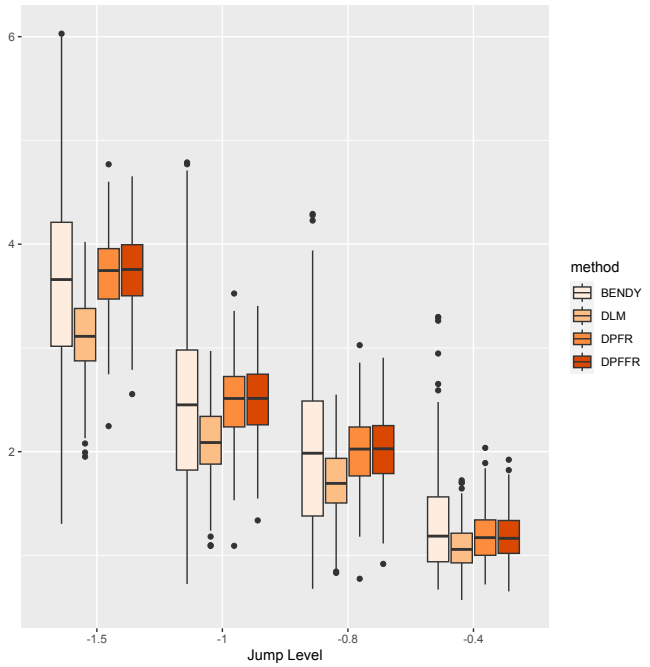

Figure 5: Distribution over 300 simulations of the average of absolute z-scores for four methods (BENDY, DLM, DPFR, DPFFR) and four jump levels (C=-1.5,-1,-0.8,-0.4).

4.1.2 Comparison with static outlier detection methods

Using the same scenario described in the previous section, we compare our methods to methods designed for static outlier detection. We have done our best to implement and interpret these methods according to how they would be used in practice. For example, for the BD method, the band depth was computed for each function within each data set. The calculated band depths were then sorted in increasing order and the function with the smallest band depth (most outlying) was assigned rank 1 while the function with the highest band depth was assigned rank 200 (because every simulated data set had $I = 200$ functions). Each simulated data set contains one simulated dynamic outlier, which receives a rank according to any method and these ranks are then accumulated over each method and simulation. Our dynamic outlier detection methods can also be used for ranking using the average, maximum,

or number of z-scores with absolute values larger than the t quantile $q_{101-59,\alpha}$.

Figure 6 provides the boxplots of the distributions of the ranks of the outlier according to various methods. Recall that lower rank is better, indicating that the outlier is given a higher importance as a potential unusual observation. Each subplot corresponds to a size of the outlier with larger outliers ($C = -1.5$) to the left and smaller outliers ($C = -0.4$) to the right. We used 12 dynamic outlier detection methods by combining the four dynamic prediction methods with three summaries (average, max, and number of absolute values of z-scores above the t quantile $q_{101-59,\alpha}$). In Figure 6 the labels are organized the same way. For example, DPFR_ave is the method DPFR using the average of the absolute z-scores and DLM_nr is the dynamic linear regression with the number of absolute values of z-scores larger than the t quantile $q_{101-59,\alpha}$. We compared our dynamic outlier detection methods with six static outlier detection methods; see Section 4.1.3 for details.

Figure 6 indicates that dynamic methods perform well at identifying the outliers among the top ranked curves. In particular, DPFR_ave, DPFFR_ave, DPFR_nr, and DPFFR_nr perform very well for $C = -1.5$ and $C = -1.0$, with DPFR_ave, and DPFFR_ave performing better for $C = -0.8$. As the outliers are getting smaller, dynamic models have a harder time identifying the outlier curve as the top likeliest curve to be different, but continue to rank it close to the top. DPFR_ave and DPFFR_ave continue to perform the best even for small deviations of the outlying curve $C = -0.4$. In contrast, all six traditional (or static methods) cannot identify the dynamic outlying curves, with the median rank being consistently close to 100, or roughly in the middle of the $I = 200$ curves. Moreover, the static methods are not sensitive to the size of the jump, whereas the dynamic outlier detection methods

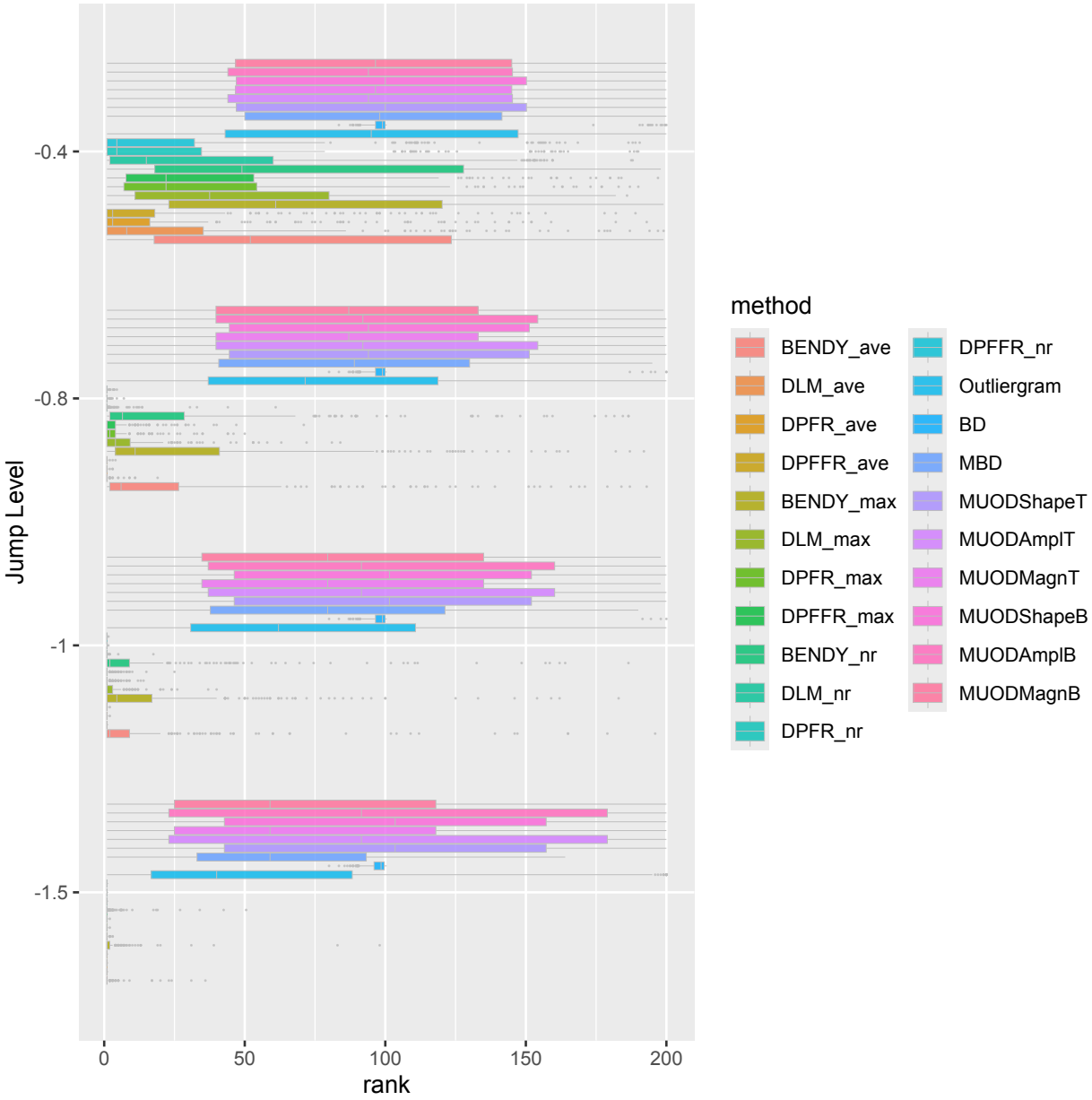

Figure 6: Simulation study results for four jump levels C for each method (organized from $C = -1.5, -1, -0.8, -0.4$ within subpanels from left to right). Each boxplot corresponds to a method and provides the distribution of the rank of the true outlier evaluated over 300 simulations.

perform better for larger jumps. This should not be surprising, given the intended use of the static outlier identification methods, but it shows that *dynamic outlier identification is a different problem than traditional functional outlier identification.*

4.1.3 Implementation strategies

This is the list of outlier detection methods in Section 4.1.2.

- Outliergram: Method proposed by Arribas-Gil and Romo (2014) implemented in the R function `roahd::outliergram` with default arguments.
- BD and MBD: The Band Depth (BD) and Modified Band Depth (MBD) methods proposed by Lopez-Pintado and Romo (2007, 2009) and implemented in the R functions `roahd::BD` and `roahd::MBD`, respectively, with default arguments.
- MUODShapeT, MUODAmplT, MUODMagnT: The massive unsupervised outlier detection (MUOD) method proposed by Azcorra et al. (2018) and implemented in the R function `fdaoutlier::muod` (Ojo et al. 2021) with the `tangent` argument. The value `indices` for `shape`, `amplitude`, and `magnitude` were obtained and provided a computed index for each case.
- BENDY_ave, BENDY_max, BENDY_nr: BENDY method of ranking based on the average, maximum, and number of absolute z-scores larger than the t quantile $q_{101-59,\alpha}$. Similar for DLM, DPFR, and DPFFR.

4.2 Simulation of type II dynamic outliers

The type I dynamic outlier consists of a constant change for one study participant after a particular point. The type II outlier consists of an incremental change/deviation

from the individual trajectory after a particular time point. More precisely, the deviation function is $L(t_m) = -0.4 \times I_{t_m \in [0.59,0.7]} - 0.8 \times I_{t_m \in (0.7,0.8]} - 1 \times I_{t_m \in (0.8,0.9]} - 1.5 \times I_{t_m \in (0.9,1.0]}$. Figure 7 displays this deviation for four simulated data sets (gray lines) with one outlier trajectory added (original trajectory shown in blue, modified trajectory shown in red). The type II dynamic outlier is relatively small when t_m close to 0.6 and increases when t_m is close to 1. All simulations and results followed the same exact procedure as in Section 4.1.

4.2.1 Results

Dynamic outlier detection methods were performed at nine time points in between the $[0, 1]$ domain. The dynamic z scores for each method were calculated starting at each of the time points $(0.59, 0.6, 0.65, \ldots, 0.95)$. At each time point where dynamic methods were performed, dynamic z scores were computed at all time points in the future, and results were summarized as in Section 4.1. Results are summarized and shown in Figure 8 across 300 simulations. As in Section 4.1, all static outlier detection methods failed to identify the type II outlier. At $0.59, 0.6, 0.65, 0.7$ all our methods consistently identify the outlier as the top outlier. However, after 0.75 the performance of methods started to degrade, though BENDY_ave, BENDY_max, BENDY_nr, DPFFR_ave, and DPFFR_max continue to perform very well, though they also fail to detect the outlier at $t_0 = 0.95$. The reason for this is that there are fewer data points $[t_0, 1]$ and the deviations between $Y_i(t)$ in $[t_0, 1]$ and $Y_i(t_0)$ are small. Moreover, for the time point 0.95, the simulated outlier has already been a dynamic outlier between time frame 0.59 and 0.95. On the other hand, when outlier detection methods were applied earlier, such as at 0.59, the dynamic outlier was detected by all dynamic methods and across all future time points as a top outlier.

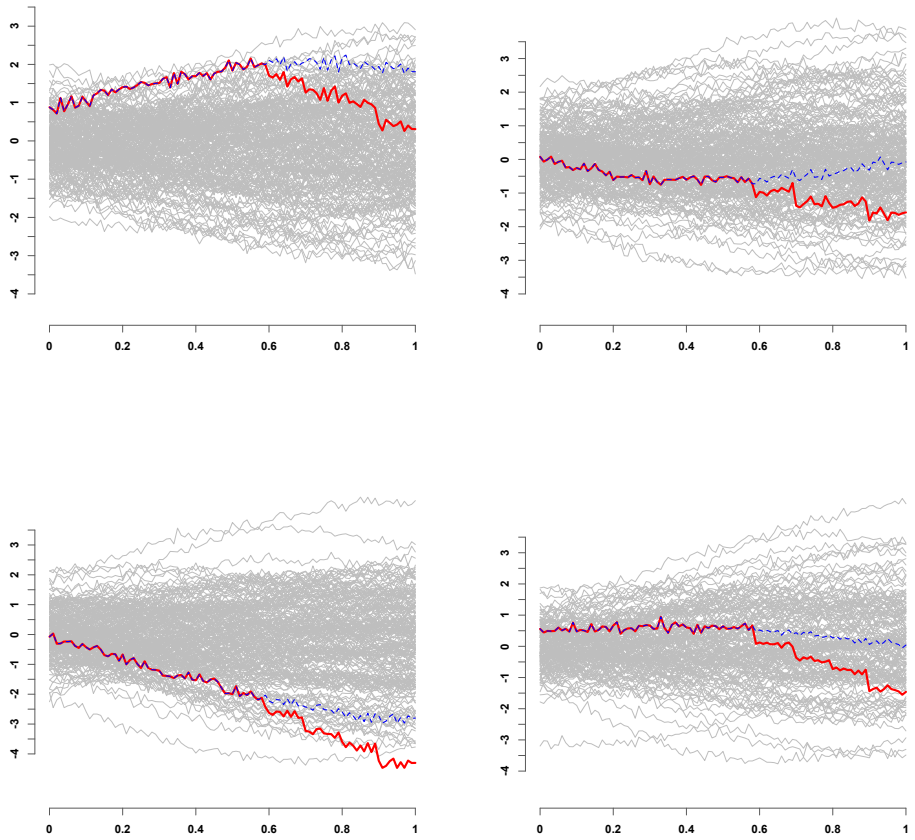

Figure 7: Simulated datasets and type II dynamic outlier. Four simulated functional datasets with $I = 200$ trajectories (light gray). The simulated dynamic outlier (red line) is obtained by adding a gradual level shift, $L(t_m)$, to the original trajectory (dashed blue line).

5 Summary

We have introduced methods designed to identify dynamic outliers in functional data. The main ideas are to: (1) use the historical information up to a given point; (2) predict the future of the trajectory based on the subject-specific historical information, population trajectories, patterns, such as time-dependent and independent covari-

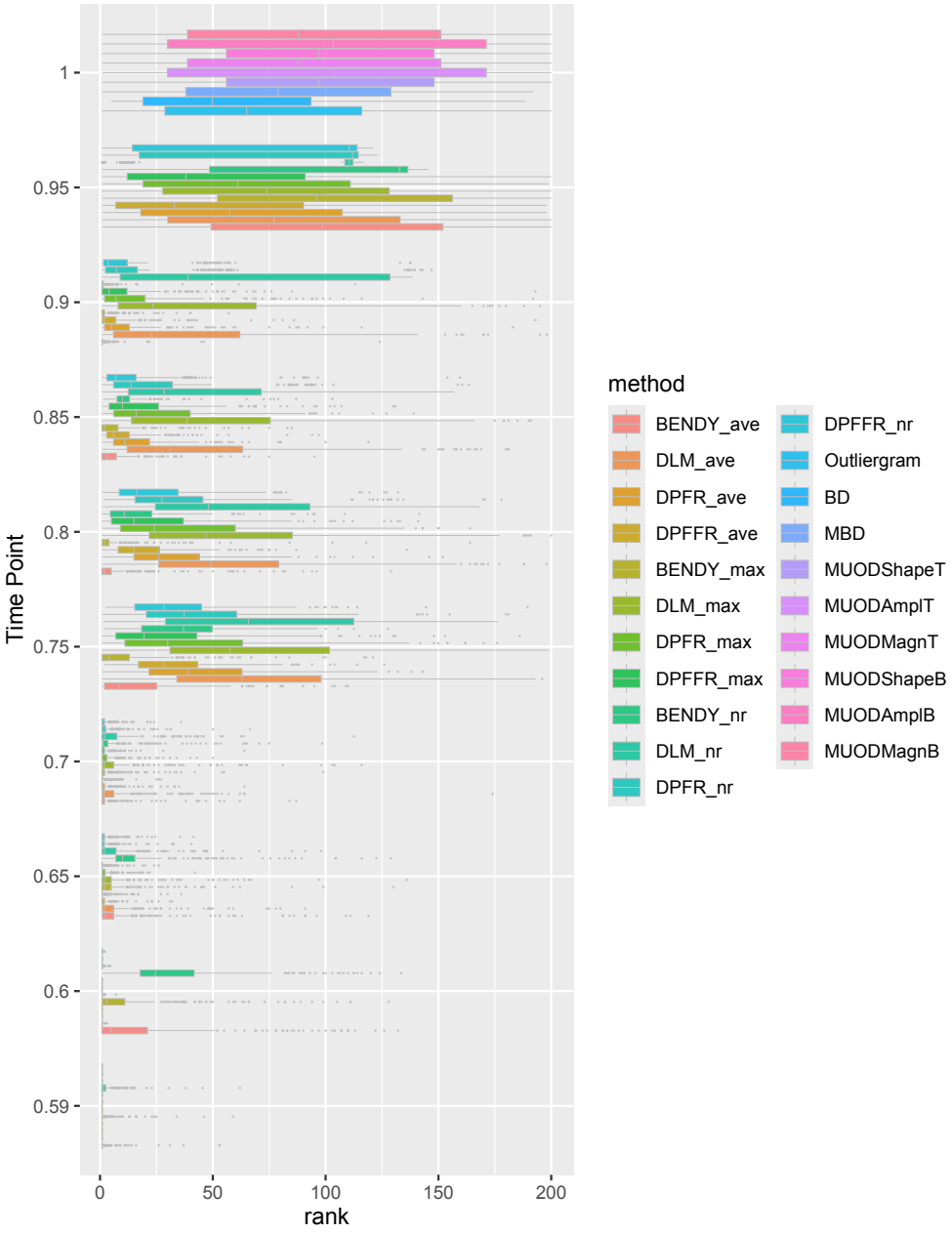

Figure 8: Results for dynamic outliers of type II.

ates; and (3) identify subject-specific outliers using the z-score for the observed versus predicted values. These methods are based on dynamic functional regression and produce outliers that are both theoretically and practically different from existent (static) methods for outlier detection on functional data. Simulations show that dy-

namic prediction identified even small dynamic outliers, while existing methods did not. In the CONTENT study application we have shown that there is very little overlap between the top outlier candidates using dynamic and static outliers. Dynamic outlier detection approaches are especially useful in the context when one is interested in early detection of abnormal patterns, such as growth patterns, which could be used for real time interventions.

Acknowledgements

William Checkley and Ciprian M. Crainiceanu were supported by a Bill & Melinda Gates Foundation Grant (OPP1114097).

References

[1] Arribas-Gil, A. and Romo, J. (2014). Shape outlier detection and visualization for functional data: the outliergram. *Biostatistics*,15(4), 603–619.

[2] Azcorra, A., Chiroque, L. F., Cuevas, R., Anta, A. F., Laniado, H., Lillo, R. E., Romo, J., and Sguera, C. (2018). Unsupervised scalable statistical method for identifying influential users in online social networks. *Scientific reports*, 8(1), 1–7.

[3] Checkley, W., Epstein, L., Gilman, R., Black, R., Cabrera, L., and Sterling, C. (1998). Effects of Cryptosporidium parvum infection in Peruvian children: growth faltering and subsequent catch-up growth. *American Journal of Epidemiology*, 148(5), 497–506.

[4] Checkley, W., Gilman, R., Black, R., Lescano, A., Cabrera, L., Taylor, D., and

Moulton, L. H. (2002). Effects of nutritional status on diarrhea in peruvian children. *Journal of Pediatrics*, 140(2), 210–218.

[5] Chiou, J.-M., Yang, Y.-F., and Chen, Y.-T. (2016). Multivariate functional linear regression and prediction. *Journal of Multivariate Analysis*, 146, 301–312.

[6] Crainiceanu, C. M., Goldsmith, J., Leroux, A., and Cui, E. (2024). *Functional data analysis with R*, Chapman & Hall/CRC, New York.

[7] Febrero, M., Galeano, P., and Gonzalez-Manteiga, W. (2015). Outlier detection in functional data by depth measures, with application to identify abnormal NOX levels. *Environmentrics*, 19, 331–345.

[8] Goldberg, Y., Ritov, Y., and Mandelbaum, A. (2014). Predicting the continuation of a function with applications to call center data. *Journal of Statistical Planning and Inference*, 147, 53–65.

[9] Goldsmith, J., Bobb, J., Crainiceanu, C. M., Caffo, B. S., and Reich, D. (2011). Penalized functional regression. *Journal of Computational and Graphical Statistics*, 20(4), 830–851.

[10] Goldsmith, J., Greven, S., and Crainiceanu, C. (2013). Corrected confidence bands for functional data using principal components. *Biometrics*, 69, 41–51.

[11] Goldsmith, J., Scheipl, F., Huang, L., Wrobel, J., Di, C., Gellar, J., Harezlak, J., McLean, M., Swihart, B., Xiao, L., Crainiceanu, C., and Reiss, P. T. (2025). `refund`: Regression with Functional Data, `http://CRAN.R-project.org/package=refund`.

[12] Grajeda, L. M., Ivanescu, A. E., Saito, M., Crainiceanu, C. M., Jaganath, D., Gilman, R. H., Crabtree, J. E., Kelleher, D., Cabrera, L., Cama, V., and Checkley,

W. (2016). Modelling subject-specific childhood growth using linear mixed-effect models with cubic regression splines. *Emerging Themes in Epidemiology*, 13, 1–13.

[13] Hyndman, R. J. and Shang, H. L. (2010). Rainbow plots, bagplots and boxplots for functional data. *Journal of Computational and Graphical Statistics*, 19(1), 29–45.

[14] Hyndman, R. J. and Shang, H. L. (2025). `ftsa`: functional time series analysis, `http://CRAN.R-project.org/package=ftsa`.

[15] Hubert, M., Rousseeuw, P. J., and Segaert, P. (2015). Multivariate functional outlier detection. *Statistical Methods and Applications*, 24, 177–202.

[16] Ieva, F., and Paganoni, A. M. (2013). Depth measures for multivariate functional data. *Communications in Statistics: Theory and Methods*, 41, 1265–1276.

[17] Ieva, F., Paganoni, A. M., Romo, J., and Tarabelloni, N. (2019). `roahd` Package: Robust analysis of high dimensional data. The R Journal, 11(2), 291–307.

[18] Ivanescu, A. E., Crainiceanu, C. M., and Checkley, W. (2017). Dynamic child growth prediction: a comparative methods approach. *Statistical Modelling*, 17(6), 468–493.

[19] Leroux, A., Xiao, L., Crainiceanu, C. M., Checkley, W. (2018). Dynamic prediction in functional concurrent regression with an application to child growth. *Statistics in Medicine*, 37(8), 1376–1388.

[20] Lopez-Pintado, S. and Romo. J. (2007). Depth-based inference for functional data. *Computational Statistics & Data Analysis*, 51, 4957-4968.

[21] Lopez-Pintado, S., and Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486), 718–734.

[22] Mejia, A. F., Nebel, M. B., Eloyan, A., Caffo, B., and Lindquist, M. A. (2017). PCA leverage: outlier detection for high-dimensional functional magnetic resonance imaging data. *Biostatistics*, 18(3), 521–536.

[23] Ojo, O., Lillo, R. E., and Fernandez Anta, A. (2021). fdaoutlier: Outlier detection tools for functional data analysis, R package, `https://CRAN.R-project.org/package=fdaoutlier`.

[24] R Development Core Team (2025). R: *A Language and Environment for Statistical Computing*, R Core Team. R Foundation for Statistical Computing, Vienna, Austria, `http://www.R-project.org`.

[25] Ren, H., Chen, N., and Zou, C. (2017). Projection-based outlier detection in functional data. *Biometrika*, 104(2), 411–423.

[26] Sawant, P., Billor, N., and Shin, H. (2012). Functional outlier detection with robust functional principal component analysis. *Computational Statistics*, 27, 83–102.

[27] Scheipl, F., Staicu, A.-M., and Greven, S. (2015). Functional additive mixed models. *Journal of Computational and Graphical Statistics*, 24(2), 477–501.

[28] Shang, H. L. (2017a). Forecasting intraday S&P 500 index returns: A functional time series approach. *Journal of Forecasting*, 36(7), 741–755.

[29] Shang, H. L. (2017b). Functional time series forecasting with dynamic updating: An application to intraday particulate matter concentration. *Econometrics and Statistics*, 1, 184–200.

[30] Wood, S. N. (2025). `mgcv`: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation, `http://CRAN.R-project.org/package=mgcv`.

[31] Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*, Chapman & Hall/CRC, New York.

[32] Zhang, W., and Wei, Y. (2015). Regression based principal component analysis for sparse functional data with applications to screening growth paths. *Annals of Applied Statistics*, 9(2), 597–620.

Appendix

Appendix material for CONTENT study data analysis

We present results for the top 10 outliers detected. Table 4 lists the ID of the outliers identified and the metric by method. For methods `moutliergram`, `multiMBD`, and `multiBD` we report the observed depth or distance for the subjects arranged in ascending order of depth, and descending order of distance.

Table 4: Results for outlier detection for the CONTENT study. Numbers placed in the column labeled "ID" represent the case numbers for the outliers detected.

`moutliergram`		`multiMBD`		`multiBD`		$\text{DPFFR}_{\text{ave}}$		$\text{DPFFR}_{\text{max}}$		DPFFR_{nr}	
ID	dist	ID	depth	ID	depth	ID	ave $\lvert z \rvert$	ID	max $\lvert z \rvert$	ID	nr
145	.24	32	.05	17	.01	12	2.25	7	4.97	12	4
115	.20	112	.06	20	.01	135	2.21	68	4.79	135	4
153	.19	76	.08	30	.01	86	2.19	98	4.27	167	4
95	.17	17	.08	32	.01	17	2.14	17	4.11	5	3
189	.16	72	.10	34	.01	188	2.03	40	4.08	40	3
132	.14	66	.10	46	.01	68	2.02	135	3.72	68	3
161	.14	25	.11	72	.01	40	1.95	132	3.58	86	3
97	.13	34	.12	75	.01	140	1.94	39	3.57	98	3
151	.13	163	.13	112	.01	196	1.82	196	3.51	140	3
63	.12	30	.13	117	.01	98	1.82	12	3.42	161	3