Pataky, T, Vanrenterghem, J, Robinson, MA and Liebl, D

On the validity of statistical parametric mapping for nonuniformly and heterogeneously smooth one-dimensional biomechanical data

On the validity of statistical parametric mapping for nonuniformly and heterogeneously smooth one-dimensional biomechanical data

Todd C. Pataky[a], Jos Vanrenterghem[b], Mark A. Robinson[c], Dominik Liebl[d]

[a]*Department of Human Health Sciences, Kyoto University, Japan*
[b]*Department of Rehabilitation Sciences, KU Leuven, Belgium*
[c]*Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, UK*
[d]*Institute for Financial Economics and Statistics, University of Bonn, Germany*

Abstract

ABSTRACT

Nonuniform (non-constant) temporal smoothness can arise in biomechanical processes like impacts, and heterogeneous smoothness (unequal smoothness across observations) can arise in mechanically diverse comparisons such as padded vs. unpadded impacts, where padded dynamics are generally smoother than unpadded dynamics. It has been reported that statistical parametric mapping's (SPM's) probability values can be invalid for such cases. The purpose of this paper was to clarify the scope of validity for SPM analysis of nonuniformly and heterogeneously smooth one-dimensional (1D) data. We simulated a variety of nonuniformly and heterogeneously smooth Gaussian 1D data over a range of smoothness values, and computed Type I error rates across 10,000 simulation iterations for each smoothness type. Results showed that, in all cases, SPM accurately controlled error at the prescribed α=0.05. Moreover, the distribution of false positives was uniform across time, implying that all regions are equally likely to produce false positives, irrespective of local roughness. We nevertheless show that cluster-level inferences (i.e., p values specific to local regions of significance) may be over-or-underestimated by approximately 0.01 (for the currently simulated scenarios), but never exceed α by definition. We conclude that SPM's null hypothesis rejection decisions are valid for both nonuniform and heterogeneous 1D data, but that clusters' p values may be marginally too small/large in rough/smooth regions, respectively. Since cluster-level p values never exceed α, these p value errors are negligible for hypothesis testing purposes. Nevertheless, inter-cluster p value comparisons should be avoided. Implications for statistical power and general results interpretation are discussed.

Corresponding Author:
Todd Pataky, pataky.todd.2m@kyoto-u.ac.jp, +81-075-751-4175

Keywords: *biomechanics; random field theory; false positives; Monte Carlo simulations; human movement analysis methods*

1. Introduction

A variety of theoretical probabilities in statistical parametric mapping (SPM) (Friston et al., 2007) require uniformly smooth continua (Cao, 1999; Cao and Worsley, 1999; Hayasaka et al., 2004; Cheng and Schwartzman, 2015), where 'uniform' smoothness implies constant smoothness across the measurement domain. While it has been shown that SPM's continuum-level inference (the inference procedure used for hypothesis testing) is valid for nonuniform smoothness (Fig.1a,b) (Hayasaka et al., 2004; Friston et al., 2007), it has also been shown that nonuniform smoothness can yield errors in other SPM probabilities, for both one-dimensional (1D) (Liebl et al., 2018) and higher-dimensional data (Hayasaka et al., 2004). One specific, recently reported problem is that naive interval-wise inference can yield highly incorrect probabilities for nonuniformly smooth data (Liebl et al., 2018), although we note that this problem is solved by SPM's region-of-interest analysis framework (Nieto-Castanon et al., 2003; Brooks et al., 2016) as implemented in open-source SPM packages like MarsBar (Brett et al., 2002) and spm1d (Pataky, 2012).

Nonuniform smoothness can arise in biomechanical systems from a variety of sources including impacts (e.g. Neptune et al., 1999; Bisseling and Hof, 2006), and high-velocity, sequential kinematics like baseball pitching (e.g. Barrentine et al., 1998) and golfing (e.g. Cahalan et al., 1991), which produce with both low- and high-frequency dynamics within single movements (Fig.1c,d). If SPM were indeed invalid for the analysis of such data it would imply that SPM must not be applied as a general analysis tool for 1D biomechanical datasets.

SPM has been extensively validated for uniform smoothness (Fig.2a), but has not, to our knowledge, been validated for nonuniform (Fig.2b) or heterogeneous smoothness (Fig.2d). We note that our smoothness terms 'uniform' and 'heterogeneous' imply stationary (constant) 1D means, and that all 1D residuals, from which SPM's probabilities are calculated, have a stationary mean of zero by definition. More generally 1D means can be nonstationary (Fig.2c), and we note that 'stationarity' is usually used in the literature to refer to both the mean and (co)variance, so 'nonstationary' 1D data are those data whose mean is non-constant and/or whose smoothness is nonuniform. The literature also uses the term 'nonisotropic smoothness' (Worsley et al., 1999), but this is specific to 2D and higher-dimensional data (Fig.3).

Nonuniform smoothness clearly exists in some 1D biomechanical datasets (Fig.1) (c.f. Cahalan et al., 1991; Barrentine et al., 1998; Neptune et al., 1999; Bisseling and Hof, 2006), but what about heterogeneously smooth data? Consider ground reaction forces (GRF) during impacts in shod vs. unshod running, for example, where impact forces have qualitatively different frequency characteristics in these two conditions (Lieberman et al., 2010). All experimental designs involving both shod and unshod running would likely contain heterogeneously smooth 1D residuals, either with constant smoothness in a single 1D residual like those depicted in Fig.2d, or with multiple frequency components in a single 1D residual (e.g. when a single individual's mean 1D shod GRF is subtracted from the same individual's mean 1D unshod GRF. SPM handles heterogeneity implicitly, using effectively just the mean first derivative (Kiebel et al., 1999), which numerically approximates the integral of the variance, thereby averaging across frequency components.

The purpose of this paper was to clarify the scope of SPM's validity for the analysis of nonuniformly and/or heterogeneously smooth one-dimensional (1D) data. Since SPM typically conducts both continuum-level and cluster-level inferences (Friston et al., 2007), where continuum-level inference pertains to the continuum maximum, and where a 'cluster' is a suprathreshold continuum region (Fig.4), we define validity separately for each type of inference as follows:

1. SPM's continuum-level inferences may be regarded as valid if they accurately control false positive rates at α=0.05, where 'accurate' is defined as numerical convergence to α over a large number of simulations involving random 1D data.
2. SPM's cluster-level inferences may be regarded as valid if their probability values are accurate.

Below we test both types of validity using a large number of numerical simulations involving nonuniformly and/or heterogeneously smooth, random 1D data. We judge validity qualitatively based on the numerical similarity between simulation-derived probabilities and SPM's theoretical probabilities. A detailed mathematical proof of the qualitative results in this paper is currently in preparation.

2. Methods

Analyses were conducted in Python 3.6.7 (van Rossum, 2014) using Anaconda 3.6.4 (Anaconda, 2018) and the open-source package spm1d (Pataky, 2012). Code for simulating nonuniform 1D data, and scripts replicating all results from this paper are available at `https://github.com/0todd0000/nonuniform1d`.

2.1. Nonuniform smoothness models

Five different nonuniform smoothness models were considered: linear, exponential, Gaussian, sigmoid step, and double sigmoid step (Fig.5). Each nonuniform model was represented as a 1D FWHM continuum (Fig.5, top panels), where 'FWHM' is the smoothness parameter from Kiebel et al. (1999) — approximately inversely proportional to the mean first derivative — and where larger local FWHM values correspond to greater local smoothness in simulated 1D data (Fig.5, bottom panels). Minimum and maximum FWHM values were set as 6.2% and 67.0%, respectively, following previously published FWHM range results for kinematic, dynamic and EMG data (Pataky et al., 2016).

2.2. Heterogeneous smoothness models

Four different heterogeneous smoothness models were tested (Fig.6). The first three were Gaussian mixtures of low- and high-frequency noise (FWHM=25% and 2%, respectively). Setting the amplitude of the high-frequency noise at three different values (0.1, 0.5 and 2.0) yielded the three models (Fig.6a-c). The fourth model was mixture of the aforementioned nonuniform sigmoid step model (Fig.5d) and high frequency noise with an amplitude of 0.1. We used these data in a paired design, as would be expected in many biomechanical experiments (e.g. in shod vs. unshod running experiments individuals typically conduct both tasks), which resulted in single 1D residuals which contained multiple frequency components, unlike the residuals depicted in Fig.2d, which have just one frequency component each.

2.3. Simulations

For each smoothness model a 10,000-iteration Monte Carlo simulation was conducted. For each iteration a smooth, random 1D dataset was generated like those depicted in Figs.5-6, with a sample

size (N) of 12, where N is the number of 1D continua. SPM's robustness to arbitrary N has been validated elsewhere (Pataky, 2016) so arbitrary N was not considered here.

For each iteration the one-sample 1D t statistic (Fig.4) was calculated and saved to disk. This resulted in 10,000 1D t statistic continua per model. All 1D continua were sampled using 101 nodes. Also for each iteration, 1D smoothness was estimated using the FWHM parameter as described in Kiebel et al. (1999). The mean FWHM value across the 10,000 iterations was used as the estimate of the model's true population level FWHM. This approach approximates the true global FWHM with negligible numerical error, and has been validated elsewhere for uniformly smooth 1D data (Pataky, 2016). We are unaware of any published FWHM estimation procedures for nonuniformly / heterogeneously smooth data. Since SPM's theoretical expectations rely on FWHM estimates, we state the *a priori* caveat that, should simulations fail to adhere to theoretical expectations, we would be unable to distinguish between (i) theoretical inconsistencies with nonuniformity / heterogeneity, and (ii) FWHM misestimation.

2.4. Analyses

The critical t statistic (t^*) at α=0.05 was calculated once per model using random field theory (RFT) (Friston et al., 2007), based on that model's simulation-estimated FWHM. An arbitrary threshold u is depicted in Fig.4; the actual critical threshold was model-dependent and in the approximate range: $3 < t^* < 4$. This and all subsequently described RFT calculations required just two parameters: (i) degrees of freedom (ν=11) and (ii) FWHM (Kiebel et al., 1999).

From each 1D t continuum the following five attributes were extracted and stored for further analysis:

- False positive $\equiv (t_{\max} > t^*)$. When the maximum t value is greater than the critical threshold the null hypothesis is rejected, implying a false positive, or an incorrect conclusion of significance.
- Position of $t_{\max}$: the position in time between 0 and 100% at which $t_{\max}$ occurred.
- Position of all false positive nodes: the positions in time of all supra-threshold nodes.

- Maximum cluster extent ($k_{\max}$) (Fig.4): the temporal width of the widest suprathreshold cluster.

- All cluster extents (k): the temporal widths of all suprathreshold clusters.

Based on these attributes we computed the overall false positive rate (FPR) in simulation as:

$$\mathrm{FPR}_{\mathrm{sim}} = \frac{\text{Number of false positives}}{\text{Number of iterations}} \tag{1}$$

An $\mathrm{FPR}_{\mathrm{sim}}$ value of α=0.05 would imply that t^*, and thus SPM, is valid for hypothesis testing. Note that $\mathrm{FPR}_{\mathrm{sim}}$ values need only be numerically close to α because α pertains to an infinite number of experiments. $\mathrm{FPR}_{\mathrm{sim}}$ values of approximately 0.04 to 0.06 suggest validity. If theoretical expectations are accurate, $\mathrm{FPR}_{\mathrm{sim}}$ values would be expected to converge to α as the number of simulation iterations increases.

We also computed the survival function $f(u) = \mathrm{P}(t_{\max} > u)$, representing the probability with which $t_{\max}$ is expected to be larger than an arbitrary threshold u. This was computed analytically using RFT, and also in simulation as:

$$f_{\mathrm{sim}}(u) = \frac{\text{Number of iterations with } (t_{\max} > u)}{\text{Number of iterations}} \tag{2}$$

Qualitative agreement between $f(u)$ and $f_{\mathrm{sim}}(u)$ would provide further evidence that SPM's RFT computations are valid. Note that the false positive rate (Eq.1) is determined from just a single point on the survival function. The critical threshold u^* is given by setting $\mathrm{P}(t_{\max} > u^*)$=$\alpha$ then solving for u^*, and exceeding this threshold implies a false positive. The purpose of considering the entire survival function is to ensure robustness to arbitrary α.

Last, we computed the cluster-level survival function $g(c) = \mathrm{P}(k_{\max} > c)$, representing the probability that the largest suprathreshold cluster's extent $k_{\max}$ is larger than c. As above, this was first computed analytically using RFT, then in simulation as:

$$g_{\mathrm{sim}}(c) = \frac{\text{Number of iterations with } (k_{\max} > c)}{\text{Number of iterations}} \tag{3}$$

As above, qualitative agreement between $g(c)$ and $g_{\text{sim}}(c)$ would provide further evidence that SPM's RFT computations are valid.

3. Results

The overall false positive rates observed in simulation (FPR_{sim}) were numerically close to α=0.05 for all nonuniform / heterogeneous smoothness models, with maximum and minimum FPR_{sim} values of 0.041 and 0.049, respectively (Figs.7-8). Probabilities of exceeding not just the critical threshold at α=0.05, but also arbitrary thresholds also closely agreed with expectation (Fig.7-8), with the possible exception of heterogeneously smooth data with high-frequency amplitude of 0.5 (Fig.8b), which showed slightly smaller probabilities than expected for p values greater than 0.1. Nevertheless, these errors were small, approximately 0.01 for p values of 0.1. These results suggest that SPM is valid for making null hypothesis rejection decisions in experiments involving nonuniformly and/or heterogeneously smooth 1D data both at α=0.05 and at arbitrary α levels, but that some type of heterogeneously smooth 1D data may diverge slightly from expected probabilities for $\alpha \geq 0.1$.

The temporal distribution of t_{max} values was approximately uniform across time for uniformly smooth data (Fig.9e) but was skewed toward the rough regions of nonuniformly smooth data (Fig.9f). This implies that t_{max} is most likely to occur in rough continuum regions. However, it does not imply that single nodes in rough continuum regions are more likely to reach significance than single nodes in smooth continuum regions, as evidenced by approximately uniform distributions across time for *all* suprathreshold nodes (not just t_{max}) (Fig.9g,h). This apparent paradox is resolved by recognizing that smooth regions are likely to yield larger suprathreshold clusters than rough regions (Fig.10f), and this balances the higher frequencies of maxima in rough regions (Fig.9f) so that the overall probability that an arbitrary node will reach significance is constant across the 1D continuum. Pointwise, this is ordinary t-testing for which smoothness is irrelevant. This implies that the exceedance probability is constant at every point.

Cluster-specific probability values agreed with theoretical (RFT) expectations for uniformly smooth data (Fig.10a) but not for nonuniformly smooth data (Fig.10b). Simulated and theoretical

results differed by approximately p=0.01 for the considered simulations/models, and cluster-level p values never exceeded α. Nonuniformly smooth data tended to yield more small clusters than uniformly smooth data (Fig.10c,d), and small clusters tended to occur most frequently in rougher regions (Fig.10e,f). The main cluster probability results (Fig.10a,b) further emphasize that SPM is valid for hypothesis testing, but also indicate that cluster-level p values may be marginally incorrect when 1D data are nonuniformly smooth. In particular, small and large clusters' p values may be overestimated and underestimated, respectively (Fig.10b).

4. Discussion

While SPM's underlying theory generally requires uniform smoothness (Cao, 1999; Cao and Worsley, 1999; Cheng and Schwartzman, 2015), this study's numerical simulations have verified previous reports that (1) SPM's continuum-level (or 'peak'-level) inference is valid for both uniformly and nonuniformly smooth continuum data (Hayasaka et al., 2004; Friston et al., 2007) (Fig.7) and that (2) SPM's cluster-level inferences may be incorrect when smoothness is nonuniform (Hayasaka et al., 2004) (Fig.10b). While nonuniform smoothness can cause the location of the maximum test statistic to shift toward rough continuum regions (Fig.9e,f), this does not imply that individual continuum nodes are more likely to reach significance in rougher regions (Fig.9h) because larger clusters are expected in smooth regions (Fig.10f).

Our simulations further suggest that SPM's continuum-level inferences appear to be valid for heterogeneously smooth continuum data (Fig.8). Nevertheless, SPM's critical thresholds may be incorrect for some types of heterogeneous data, particularly those involving different-frequency noise components with similar amplitudes (Figs.6,8b), but only when α is relatively large ($\alpha \geq 0.1$). We are unaware of any previous investigations of heterogeneous smoothness in the context of SPM analysis, so are unable to compare these results to the literature.

Last, our results suggest that cluster-level probability errors are likely to be small, approximately 0.01 (Fig.10b), even with relatively pronounced nonuniformity (Fig.5d). Moreover these probability values never exceed α, by definition (Friston et al., 1994).

Since the null hypothesis rejection decision is governed only by continuum-level inference, these

results suggest that SPM appears to be valid for testing hypotheses regarding 1D data at α=0.05, even when those data have arbitrary smoothness characteristics. Nevertheless, cluster-level probabilities may have errors of approximately 0.01, without exceeding α, so caution should be exercised for between-cluster and between-dataset comparisons.

All aforementioned results pertain only to SPM's parametric (i.e., analytical) probabilities. A simple way to avoid smoothness-related problems is to instead use non-parametric inference (Nichols and Holmes, 2002), an approach which has been shown to yield qualitatively identical results to parametric inference for a variety of biomechanical datasets (Pataky et al., 2015; Warmenhoven et al., 2018). This type of nonparametric inference does not rely on smoothness estimates, instead incorporating smoothness implicitly into its data-based, empirical distributions from which probabilities are calculated, and this approach has also been advocated in functional data analysis (Ramsay and Silverman, 2005) as a useful way to conduct hypothesis testing on 1D continuum data. Parametric inference is nevertheless much faster, requiring only approximately 50 ms; nonparametric inference requires 5 s or more, depending on the size of the dataset. Since speed should not be the motivating factor for choosing an inference method, we recommend regularly conducting both parametric and nonparametric analysis using SPM. Insofar as the results are qualitatively similar, the parametric approach's assumptions of normality are likely reasonable.

A previous study (Liebl et al., 2018, Fig.1) reported similar results to this paper's Fig.9f, and pointed to the potential problem that a simple interval-wise inference can lead to incorrect inferences. These results pertain to the theoretical results that, if a single continuum-wide threshold is adopted for hypothesis testing, AND multiple regions are subsequently analyzed independently, rougher continuum regions can yield false positives at a rate greater than α. However, this type of separate-region analysis, which is termed 'region of interest' (ROI) analysis in the SPM literarture (Nieto-Castanon et al., 2003; Brooks et al., 2016), requires that a separate threshold be computed for each ROI, precisely because different ROIs can exhibit different smoothness characteristics. In other words, SPM's continuum-level inferences are also valid when applied to individual ROIs, because each ROI is itself regarded as an independent continuum. Thus the result from Liebl et al. (2018) implies only that an ROI implementation which exploits a uniform smoothness hypothesis

can produce false positive rates deviations from α if smoothness is only estimated once, for the whole continuum, and the smoothness is nonuniform across the continuum. This problem can be avoided using software packages that support explicit ROI analyses (Brett et al., 2002; Pataky, 2012).

The primary limitation of this study was that (mis)registration was not considered, where registration refers to continuum alignment (e.g., temporal alignment). The smoothness models in Fig.5 assume that the locations of smoothness transitions are constant across continua. For example, the sigmoid step model (Fig.5d) assumes that the smoothness changes relatively abruptly in all continua at a position of time = 50%. For real biomechanical datasets registration is usually achieved through linear interpolation (i.e., interpolating linearly between 'start'=0 and 'stop'=100%). However, this type of registration may be too simple for some datasets whose temporal events remain misaligned following linear registration, in which case nonlinear registration may be more appropriate (Ramsay and Li, 1998; Sadeghi et al., 2000; Moudy et al., 2018). The choice of linear vs. nonlinear registration, and more generally of how aligned the data should be, is not straightforward. Linear registration retains relative timing differences, and these timing differences may be important for interpreting the underlying dynamics, for example in high velocity tasks (Barrentine et al., 1998; Cahalan et al., 1991) where the relative timing between movements might be able to explain velocity outcome differences. Nonlinear registration could destroy this relative timing information. We therefore regard registration as necessary for, but independent of SPM analysis. We can recommend only that researchers try different registration procedures in a sensitivity analysis, and to explain any differences that emerge.

A secondary limitation is that we are unaware of any published random number generators for nonuniformly smooth 1D data. This study adapted a previously published random number generator for *uniformly* smooth 1D data (Pataky, 2016) (see code in this study's repository; link available at beginning of the Methods section). While these data are Gaussian, so are expected to adhere to expected characteristics of Gaussian random fields (as demonstrated in Fig.7), they are not standard Gaussian (i.e., with a mean and variance of zero and one, respectively) so we caution against their general use.

In summary, this study has clarified that SPM's null hypothesis rejection decisions at α=0.05 are robust to both nonuniform and heterogeneous smoothness in 1D datasets, and thus that SPM appears to be valid for 1D data analysis, even when those data have arbitrary smoothness characteristics. However, p values associated with small and large clusters may be over- and under-estimated, respectfully. These errors were approximately 0.01 for the considered simulations, and less than α by definition, and do not affect the null hypothesis rejection decision, so can be regarded as negligible for the purposes of hypothesis testing. When data are nonuniformly or hetereogeneously smooth caution should be employed when interpreting cluster-level probabilities.

Acknowledgments

This work was supported in part by Kiban B Grant 17H02151 from the Japan Society for the Promotion of Science. The authors wish to thank Chris Richter for alerting us to this issue, and for bringing together this paper's co-authors.

Conflict of Interest Statement

The authors report no conflict of interest, financial or otherwise.

References

Anaconda, 2018. Anaconda Software Distribution version 3-5.1. URL: `https://anaconda.com`.

Barrentine, S. W., Matsuo, T., Escanilla, R. F., Fleisig, G. S., and Andrews, J. R., 1998. Kinematic analysis of the wrist and forearm during baseball pitching. *Journal of Applied Biomechanics 14*, 24–39.

Bisseling, R. W., and Hof, A. L., 2006. Handling of impact forces in inverse dynamics. *Journal of Biomechanics 39*, 2438–2444.

Brett, M., Anton, J. L., Valabregue, R., and Poline, J. B., 2002. Region of interest analysis using the MarsBar toolbox for SPM 99. *NeuroImage 16*, S497.

Brooks, J. L., Zoumpoulaki, A., and Bowman, H., 2016. Data-driven region-of-interest selection without inflating Type I error rate. *Psychophysiology 54*, 100–113.

Cahalan, T. D., Cooney, W. P., Tamai, K., and Chao, E. Y., 1991. Biomechanics of the golf swing in players with pathologic conditions of the forearm, wrist, and hand. *The American Journal of Sports Medicine 19*, 288–293.

Cao, J., 1999. The size of the connected components of excursion sets of χ 2, t and F fields. *Advances in Applied Probability 31*, 579–595.

Cao, J., and Worsley, K., 1999. The geometry of correlation fields with an application to functional connectivity of the brain. *The Annals of Applied Probability 9*, 1021–1057.

Cheng, D., and Schwartzman, A., 2015. Distribution of the height of local maxima of Gaussian random fields. *Extremes 18*, 213–240.

Friston, K., Worsley, K., Frackowiak, R., Mazziotta, J., and Evans, A., 1994. Assessing the Significance of Focal Activations Using Their Spatial Extent. *Human Brain Mapping 1*, 210–220.

Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., and Penny, W. D., 2007. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. London: Elsevier.

Hayasaka, S., Phan, K. L., Liberzon, I., Worsley, K. J., and Nichols, T. E., 2004. Nonstationary cluster-size inference with random field and permutation methods. *NeuroImage 22*, 676–687.

Kiebel, S., Poline, J., Friston, K., Holmes, A., and Worsley, K., 1999. Robust Smoothness Estimation in Statistical Parametric Maps Using Standardized Residuals from the General Linear Model. *NeuroImage 10*, 756–766.

Lieberman, D. E., Venkadesan, M., Werbel, W. A., Daoud, A. I., D'andrea, S., Davis, I. S., Mang'eni, R. O., and Pitsiladis, Y., 2010. Foot strike patterns and collision forces in habitually barefoot versus shod runners. *Nature 463*, 531–535.

Liebl, D., Hamacher, D., and Zech, A., 2018. A cautionary note on the use of one dimensional statistical parametric mapping in Biomechanics. In M. Lames, D. Link, and V. Senner (Eds.), *Sportinformatik und Sporttechnologie* (pp. 39–40). München.

Moudy, S., Richter, C., and Strike, S., 2018. Landmark registering waveform data improves the ability to predict performance measures. *Journal of Biomechanics 78*, 109–117.

Neptune, R. R., Wright, I. C., and van den Bogert, A. J., 1999. Muscle coordination and function during cutting movements. *Medicine & Science in Sports & Exercise 31*, 294–302.

Nichols, T., and Holmes, A., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping 15*, 1–25.

Nieto-Castanon, A., Ghosh, S. S., Tourville, J. A., and Guenther, F. H., 2003. Region of interest based analysis of functional imaging data. *NeuroImage 19*, 1303–1316.

Pataky, T. C., 2012. One-dimensional statistical parametric mapping in Python. *Computer Methods in Biomechanics and Biomedical Engineering 15*, 295–301.

Pataky, T. C., 2016. rft1d: Smooth One-Dimensional Random Field Upcrossing Probabilities in Python. *Journal of Statistical Software 71*, 1–22.

Pataky, T. C., Vanrenterghem, J., and Robinson, M. A., 2015. Zero- vs. one-dimensional, parametric vs. non-parametric, and confidence interval vs. hypothesis testing procedures in one-dimensional biomechanical trajectory analysis. *Journal of Biomechanics 48*, 1277–1285.

Pataky, T. C., Vanrenterghem, J., and Robinson, M. A., 2016. The probability of false positives in zero-dimensional analyses of one-dimensional kinematic, force and EMG trajectories. *Journal of Biomechanics 49*, 1468–1476.

Ramsay, J. O., and Li, X., 1998. Curve registration. *Journal of the Royal Statistical Society Series B 60*, 351–363.

Ramsay, J. O., and Silverman, B. W., 2005. *Functional Data Analysis*. New York: Springer-Verlag.

van Rossum, G., 2014. The python library reference release 3.7.1. URL: `https://docs.python.org/3/library/`.

Sadeghi, H., Allard, P., Shafie, K., Mathieu, P., Sadeghi, S., Prince, F., and Ramsay, J., 2000. Reduction of gait data variability using curve registration. *Gait and Posture 12*, 257–264.

Warmenhoven, J., Harrison, A., Robinson, M. A., Vanrenterghem, J., Bargary, N., Smith, R., Cobley, S., Draper, C., Donnelly, C., and Pataky, T., 2018. A force profile analysis comparison between functional data analysis, statistical parametric mapping and statistical non-parametric mapping in on-water single sculling. *Journal of Science and Medicine in Sport 21*, 1100–1105.

Worsley, K., Andermann, M., Koulis, T., MacDonald, D., and Evans, A., 1999. Detecting changes in nonisotropic images. *Human Brain Mapping 8*, 98–101.

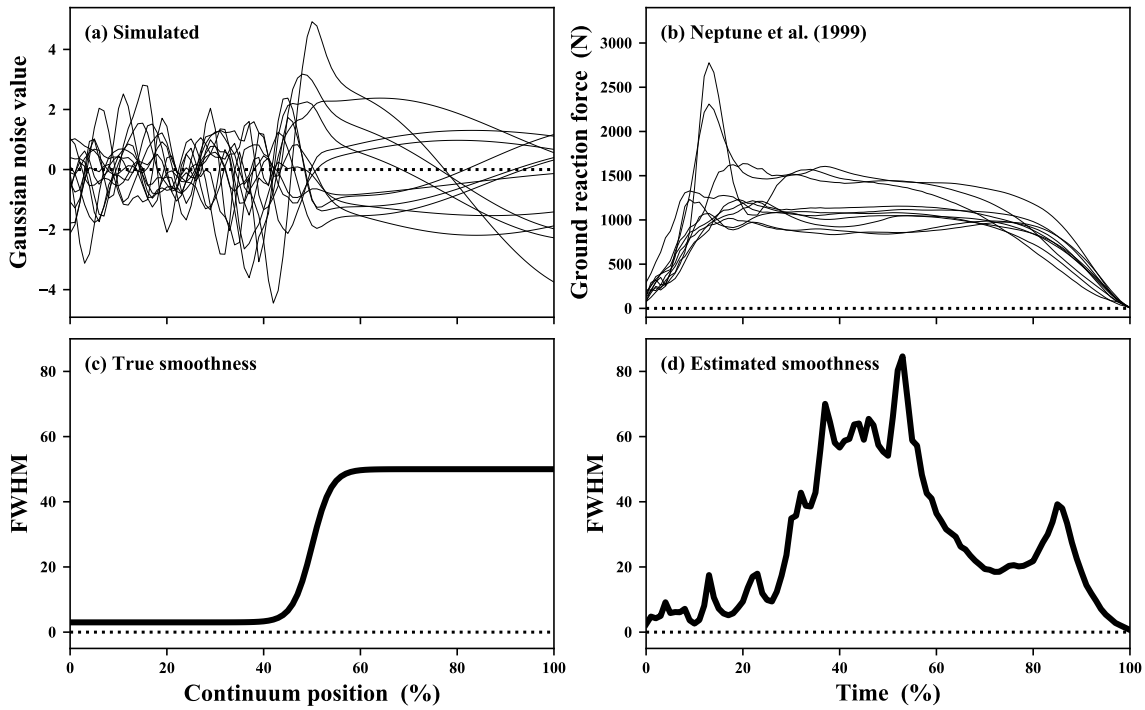

Figure 1: Example nonuniform 1D smoothness. (a) Simulated nonuniformly smooth 1D Gaussian continua; the first 40% and final 40% of the continuum have qualitatively different smoothness. (b) Vertical ground reaction force data during side-cutting, within-participant averages from Neptune et al. (1999). (c) True smoothness of the simulated data in (a), with a sigmoidal transition between initial and final smoothness values over the central 20% of the continuum; here 'FWHM' is the full-width-at-half-maximum of a Gaussian kernel which was convolved with uncorrelated (perfectly rough) Gaussian 1D data to yield the data in (a). (d) Estimated FWHM for the data in (b); estimated using residual 1D continua (i.e. mean-subtracted continua) following Kiebel et al. (1999).

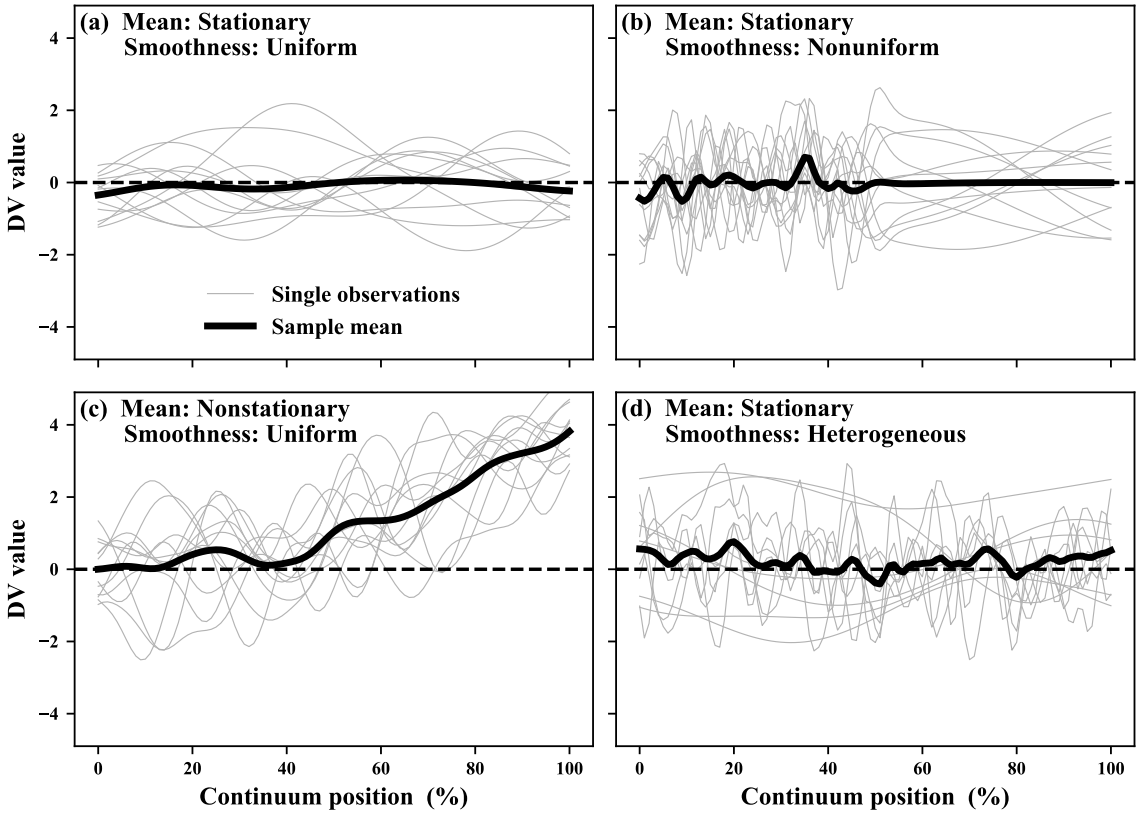

Figure 2: 1D smoothness terminology. DV = dependent variable; note that 1D 'continuum position' (e.g. time) is the measurement domain and not an independent variable. (a) 'Stationary' implies a constant population-level 1D mean. The depicted 1D mean is non-constant because it is the sample mean; an infinitely large sample would yield the expected constant population mean of zero. 'Uniform' implies constant smoothness. (b) 'Nonuniform' implies position-dependent smoothness. (c) 'Nonstationary' implies a position-dependent mean. (d) 'Heterogeneous' implies different smoothness characteristics amongst the 1D residuals in a given dataset. Note that nonstationarity (c) is irrelevant to SPM (see text).

.

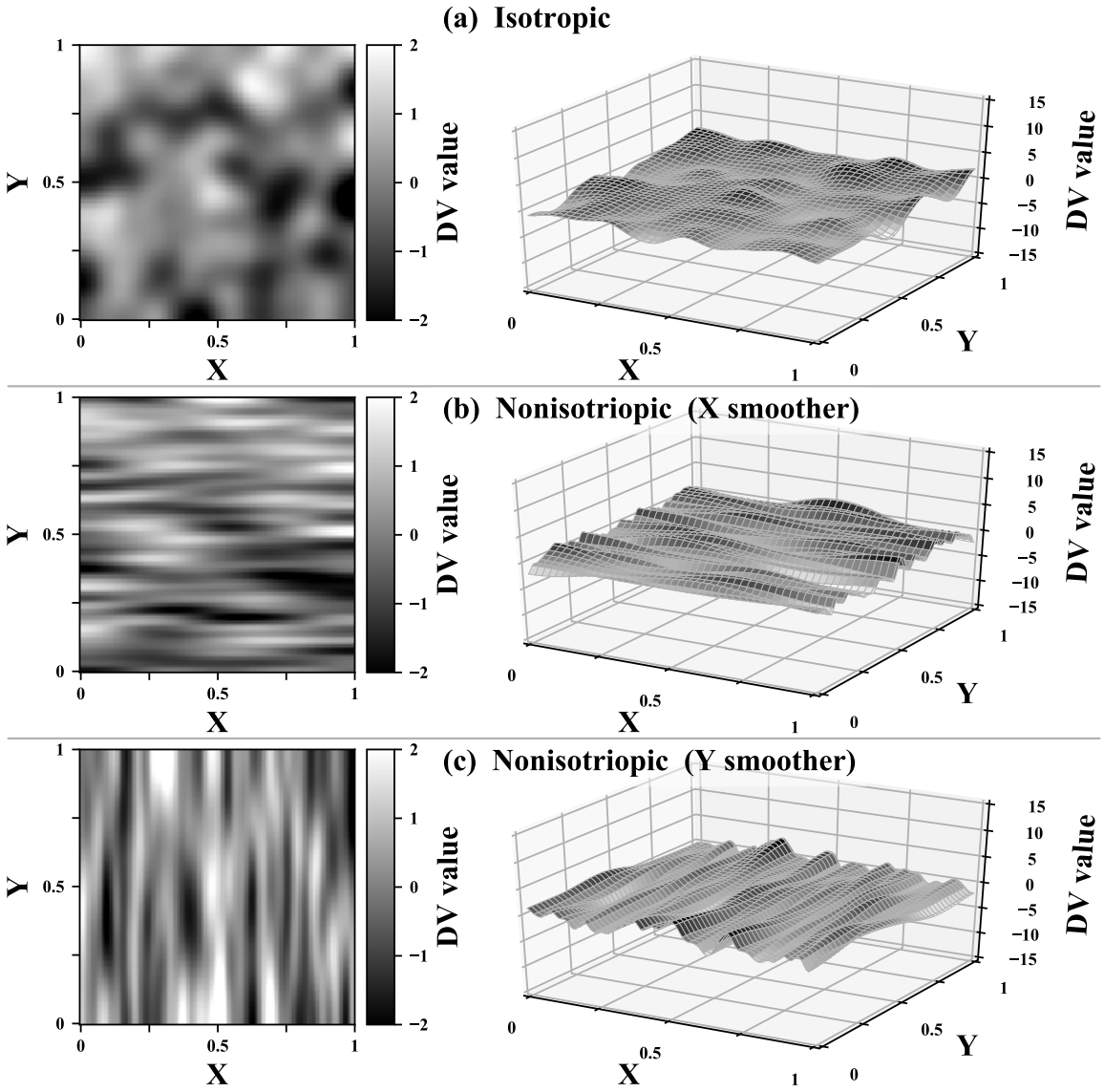

Figure 3: 2D smoothness terminology (not directly relevant to this paper, but important for distinguishing this paper's 'nonuniform smoothness' from the literature's 'nonisotropic smoothness'). DV = dependent variable. Left and right panels depict 2D continua as 2D images and 3D surfaces, respectively. (a) 'Isotropic' implies equivalent smoothness in the X and Y directions. (b,c) 'Nonisotropic' smoothness implies direction-dependent smoothness; panels (b) and (c) depict 2D continua with greater smoothness in the X and Y directions, respectively. Nonisotropic smoothness is irrelevant to 1D data but is presented here for terminology completeness.

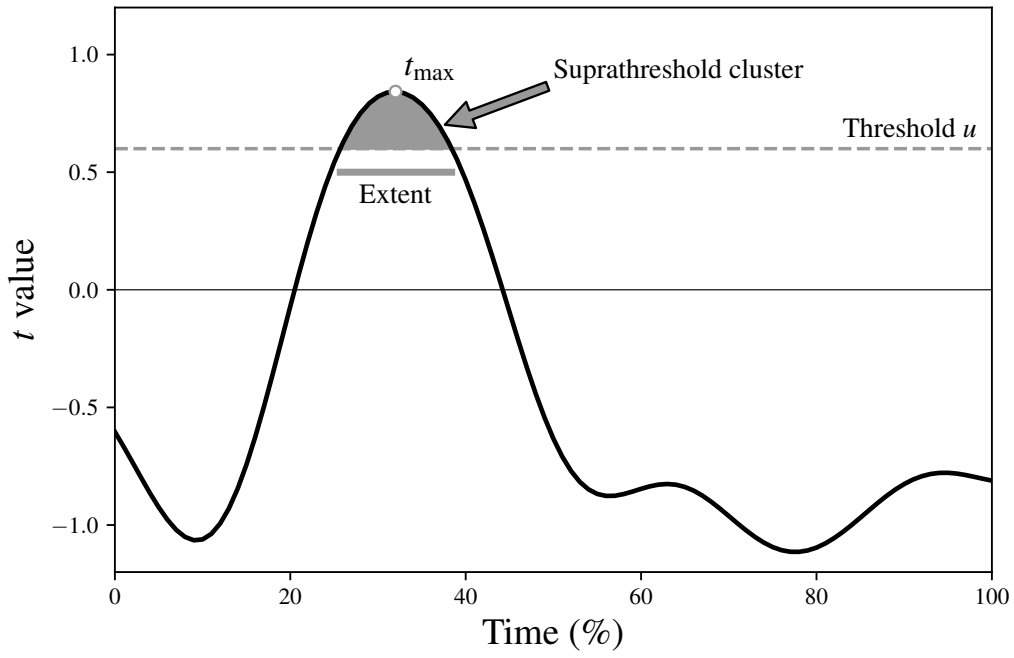

Figure 4: Example suprathreshold cluster. Here an example 1D t continuum has been thresholded at a value u to yield one suprathreshold cluster, which can be characterized by both its maximum value ($t_{\max}$) and its temporal extent.

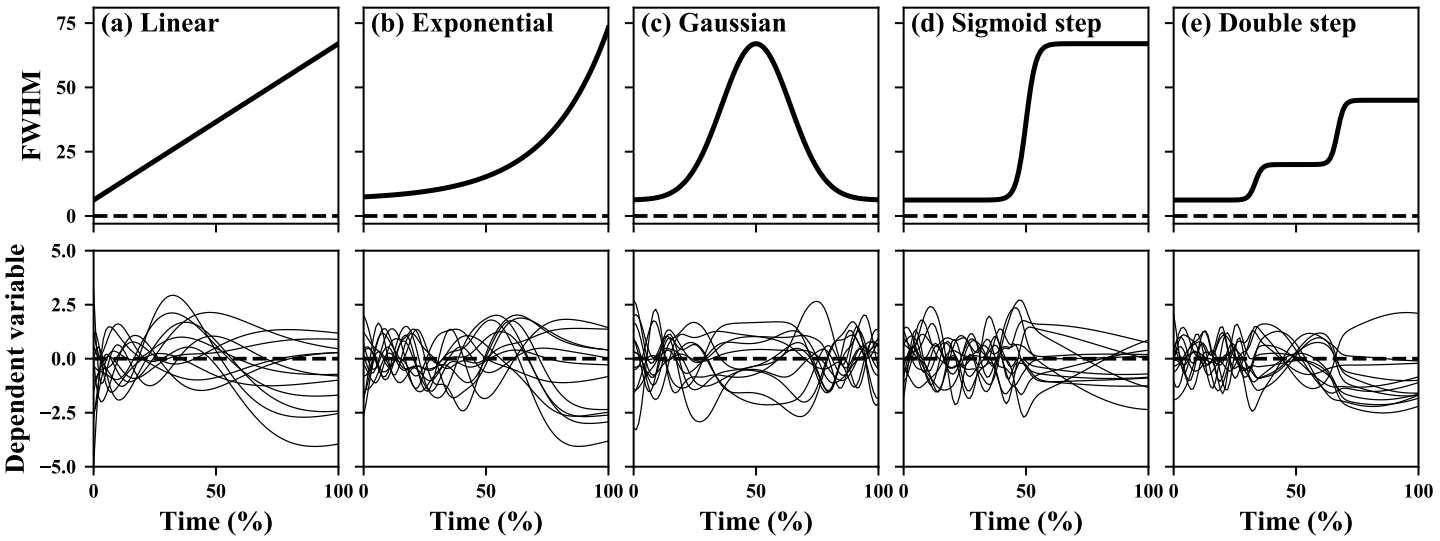

Figure 5: Nonuniformity models. The smoothness parameter FWHM is the full width at half maximum of a Gaussian smoothing kernel (see Kiebel et al. 1999). Top panels: smoothness profiles. Bottom panels: resulting random, nonuniformly smooth continua, 12 continua shown per panel.

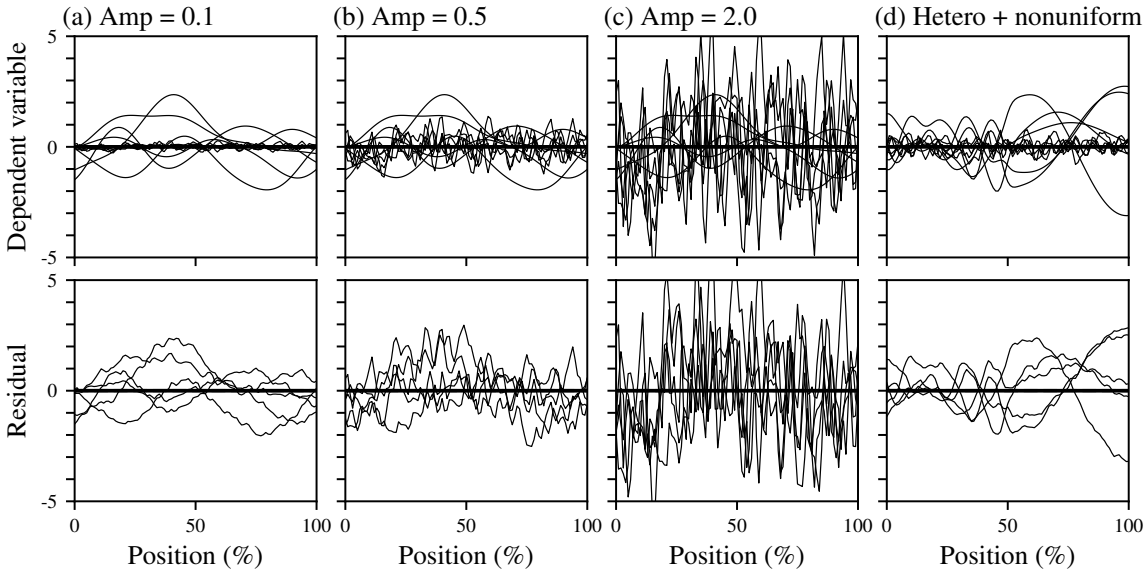

Figure 6: Heterogeneity models. (a-c) Top panels: random datasets containing five 1D observations for each of high frequency (FWHM=2) and low frequency noise (FWHM=25); data in these three panels are identical, except the high-frequency noise has been amplified by the shown *Amp* values. Bottom panels: residuals computed from the top panel data by pairing each high frequency noise observation with one low frequency observation. (d) Top panel: hybrid noise model containing observations of both nonuniform sigmoid noise (Fig.5d) and high-frequency noise (FWHM=2). Bottom panel: residuals formed by pairing nonuniform with high frequency noise observations. Only the residuals (bottom panels) were analyzed because only the residuals have controlled variance across observations.

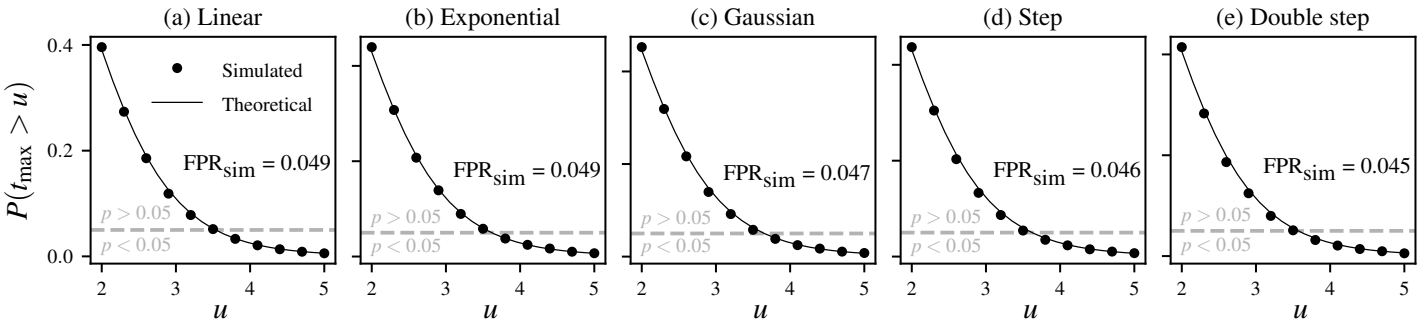

Figure 7: Survival functions (Eq.2) representing the probability that the 1D t continuum's maximum value ($t_{\max}$) exceeds an arbitrary threshold u. Each panel represents one of the five nonuinform smoothness models from Fig.5. For each model, 10,000 simulation iterations were conducted with 12 random continua generated for each iteration. Overall false positive rates observed in simulation ($\mathrm{FPR}_{\mathrm{sim}}$) are displayed. The critical threshold used for hypothesis testing was α=0.05, depicted as a grey hashed line in each panel.

.

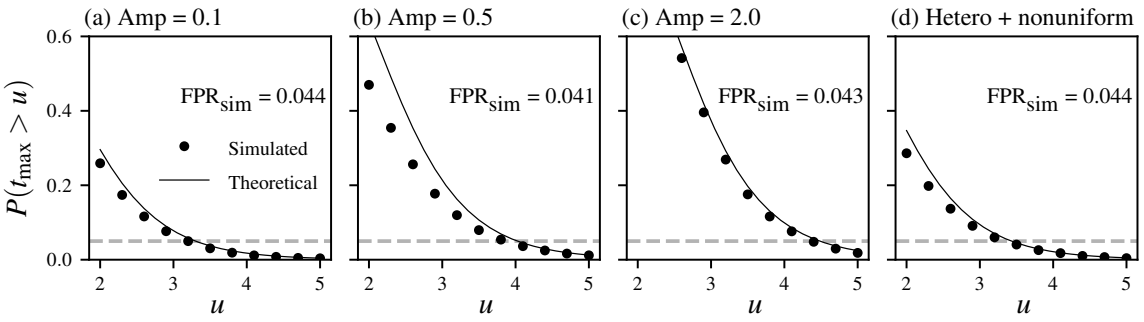

Figure 8: Survival functions (Eq.1) for heterogeneous 1D data (see Fig.6). Data presented as in Fig.7.

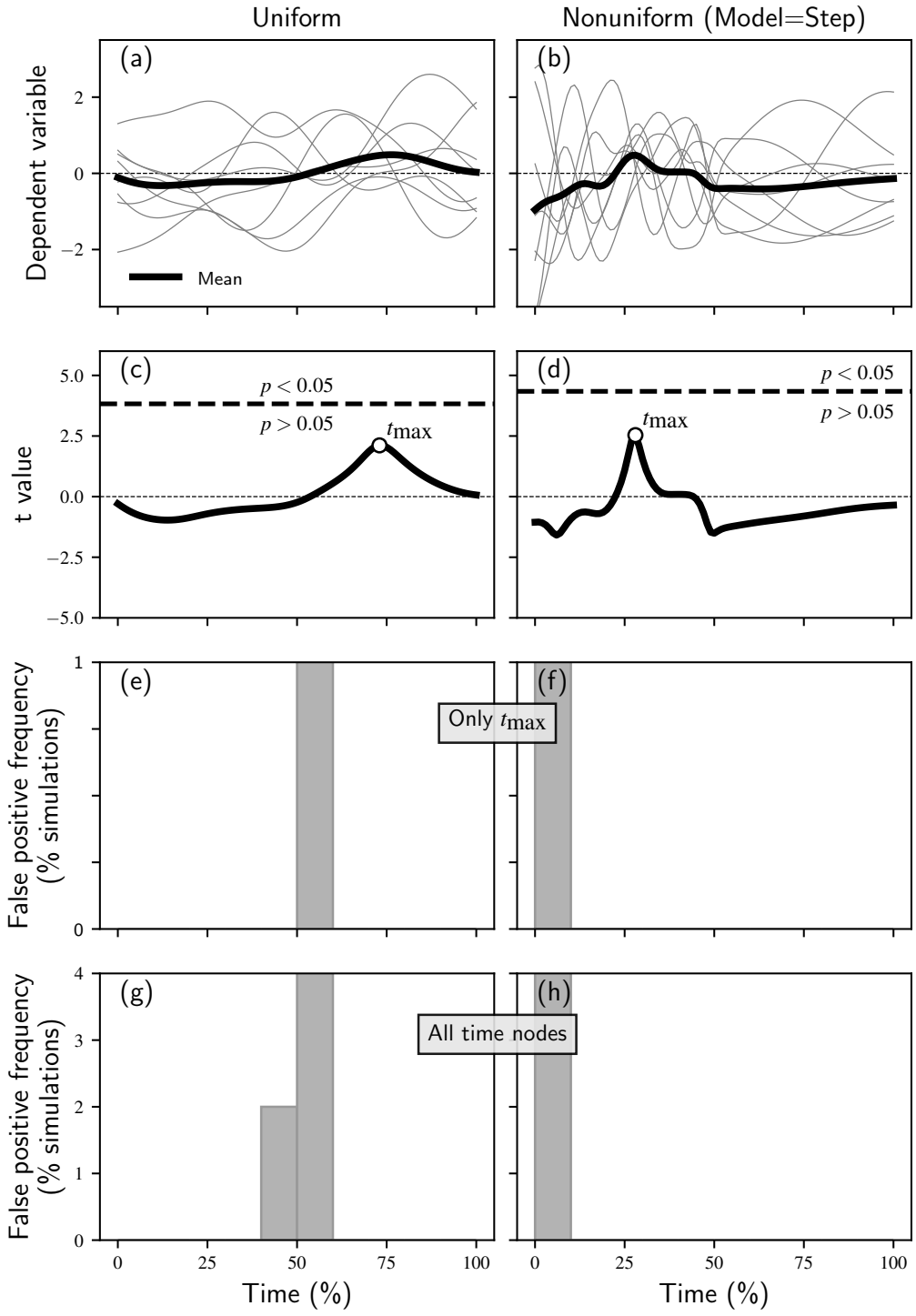

Figure 9: Effects of nonuniform smoothness on temporal distribution of false positives. Left and right panels depict uniform and nonuniform smoothness, respectively. (a,b) Example random 1D data with 1D means depicted. (c,d) SPM results depicting t continua, maximum t values ($t_{\max}$) and critical thresholds (hashed lines). (e,f) Temporal distribution of false positives for only $t_{\max}$; a bar with a value of 0.3%, for example, implies that 0.3% of all simulations (an average of 3 in 1000 simulations) yielded a false positive in the bar's temporal window. (g,h) Temporal distribution of false positives for all time nodes.

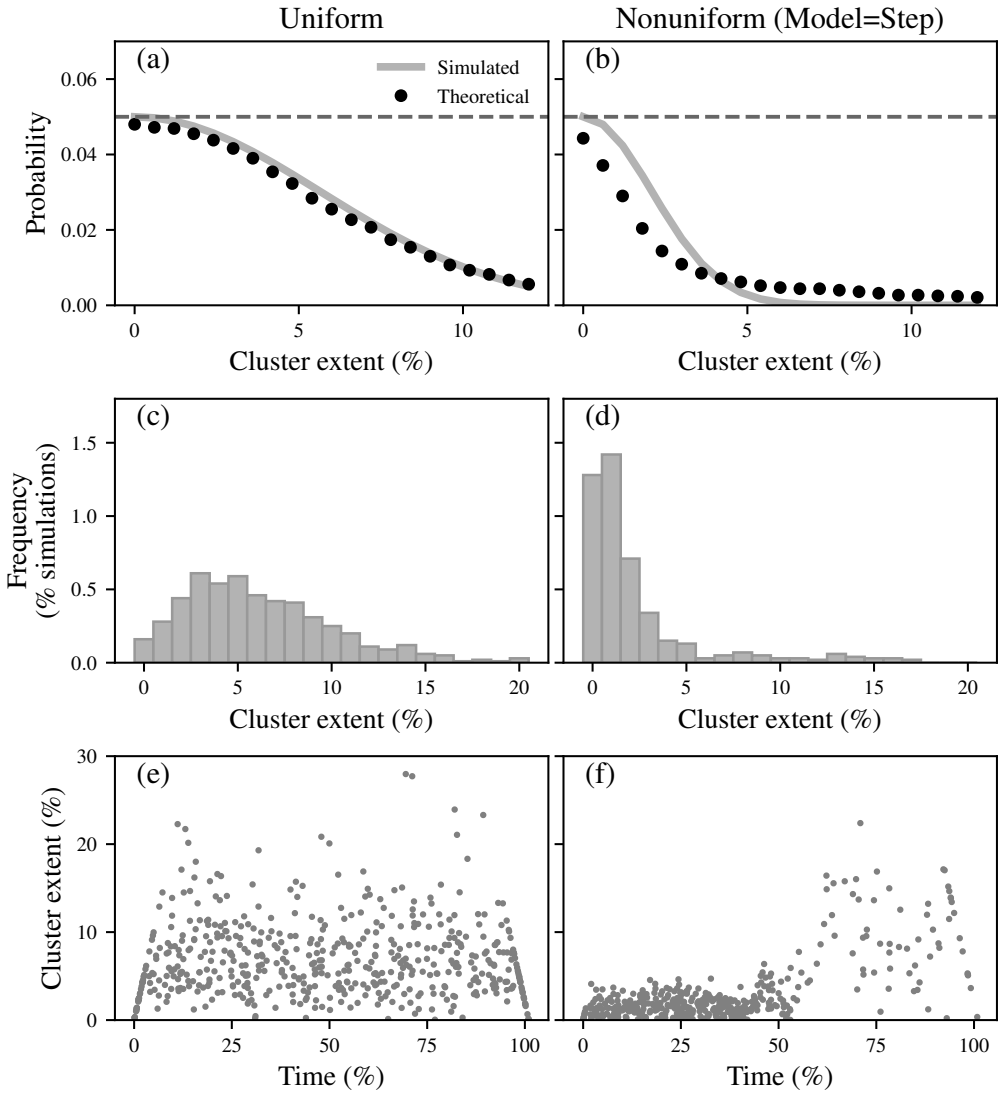

Figure 10: Effects of nonuniform smoothness on cluster extents (see Fig.4 for a description of 'cluster extent'). Left and right panels depict uniform and nonuniform smoothness, respectively. (a,b) Probability that smooth 1D random data will yield a 1D t continuum with the given cluster extent; an infinitely small cluster (extent=0) has a probability of α=0.05 by definition. (c,d) Distribution of cluster extents observed in simulations. (e,f) Scatterplot depicting where in time clusters of the given extent were observed. Note that a cluster with extent k can only occur when centered at a point within the range: $[\frac{k}{2},\ (1-\frac{k}{2})]$; a cluster with extent=12%, for example, can only occur in the time range [6%, 94%].

All supplementary material associated with this manuscript are available in this project's public GitHub repository:

https://github.com/0todd0000/nonuniform1d