

## Appendix B. SPM solution to the multiple comparisons problem

When performing  $N$  statistical tests, the probability that at least one of those tests will reach an arbitrary threshold is greater than the probability that one particular test will reach the same threshold. This is known as the multiple comparisons problem. This Appendix describes how SPM handles multiple comparisons across the whole trajectory.

Since SPM results contain a test statistic value at each point in the trajectory (Fig.2, main manuscript), one may be tempted to infer that a separate statistical test has been conducted at each node. In fact, SPM conducts just one test. To appreciate this fact, first note that a test statistic value is just a statistic, similar to descriptive statistics like the mean and standard deviation. The only reason it is called a ‘test statistic’ is because its behavior under the null hypothesis is mathematically well-defined, so we are able to use it as the basis for hypothesis testing. Simply computing the value of a test statistic does not constitute hypothesis testing. Thus computing test statistic values at each point in the trajectory constitutes zero hypothesis tests.

To further appreciate the nature of SPM’s single hypothesis test, consider the differences between rough and smooth trajectories (Fig.S2). Imagine ten rough (uncorrelated) trajectories like those in Fig.S2a, five for each of two groups, and imagine that we wish to test the null hypothesis that the groups’ mean trajectories are not different. In this case we would have to

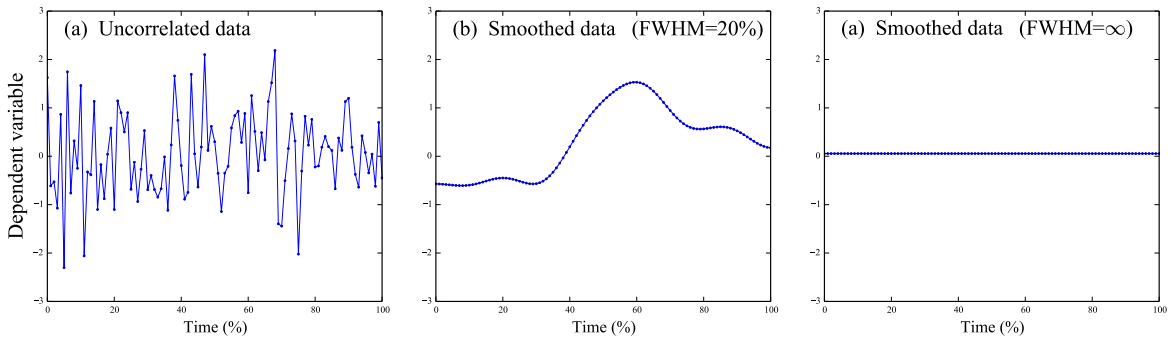


Figure S2: Gaussian random data. (a) Uncorrelated Gaussian data. (b) The same data as (a), but smoothed by a Gaussian kernel with a full-width at half-maximum (FWHM) of 20%. (c) The same data as (a), but with infinite smoothness.

adopt a correction for multiple comparisons. An obvious choice is the Šidák correction, which ensures a family-wise error rate of  $\alpha$  across all time nodes. If our trajectory contains 101 nodes, then an equivalent interpretation is that each trajectory embodies 101 independent ‘processes’.

Next consider the opposite extreme: infinite smoothness (Fig.S2c). Infinitely smooth trajectories are equivalent to single scalars in the sense that each infinitely smooth trajectory response embodies just one ‘process’. In this case we do not need a correction for multiple comparisons because there is just one ‘process’ to test.

SPM deals with the intermediate situation (Fig.S2b). When the data are smooth, but not infinitely smooth, the number of ‘processes’ is greater than one and less than 101. The actual number can be estimated from temporal gradients: large gradients imply rapid changes and therefore many processes, and small gradients imply the opposite. In the SPM literature these ‘processes’ are referred to as ‘resolution elements’ or ‘resels’, and robust techniques exist for estimating resel counts in smooth  $n$ D data.<sup>2</sup>

One appealing feature of SPM’s resel approach to the multiple comparisons problem is that it is robust to changes in sampling frequency.<sup>3</sup> We could sample the data in Fig.S2b at 1000 Hz, or even 1 GHz, but the resel count would not change because temporal gradients measure signal change per unit time, not change per time node. Moreover, the proportion of the trajectory that exceeds an arbitrary threshold is also very stable to sampling rate differences. The appealing feature is thus that the resel count, and ultimately the SPM result, is defined by the experimental trajectory itself, and not by our arbitrary sampling of that trajectory.

SPM’s resel correction for multiple comparisons is thus more accurate than a Šidák correction. A Šidák correction indeed reduces the Type I error rate, but far beyond the intended target of  $\alpha$ . It can therefore greatly *increase* the Type II error rate: an overly-severe threshold impairs our ability to detect true signal. With resel count estimates, SPM provides tighter control over both Type I and Type II error.

Last, it is interesting to note that the Šidák and SPM thresholds converge when trajec-

---

<sup>2</sup>Kiebel S, Poline J, Friston K, Holmes A, Worsley K (1999). Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. *NeuroImage* 10(6): 756–766.

<sup>3</sup>Provided the data are sampled above the Nyquist frequency.

tories become very rough (Friston et al. 2007). The reason is that gradient estimation, which underlies resel count estimation, become increasingly erratic and less accurate as trajectories approach a completely uncorrelated state (Fig.S2a). The precise point of this convergence is analytically derivable, and is typically incorporated into SPM inference procedures to ensure robustness to very rough continua.