

# Non-Parametric Combination for Analyses of Multi-Modal Imaging

<http://brainer.org>Anderson M. Winkler<sup>a,b,c</sup>, Stephen M. Smith<sup>a</sup>, Thomas E. Nichols<sup>a,d</sup>

(a) Oxford Centre for Functional MRI of the Brain, University of Oxford, UK; (b) Global Imaging Unit, GlaxoSmithKline, London, UK; (c) Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA; (d) Department of Statistics &amp; Warwick Manufacturing Group, University of Warwick, Coventry, UK.

## 1 Introduction

Inference that involves more than one imaging modality has potential to answer complex questions and to increase power. Likewise, jointly analysing a set of images from one modality, e.g. multiple task-free brain networks (a.k.a., resting state networks, RSNs), can improve interpretability and power.

Whether multiple modalities or multiple images from one modality, this combination complicates the statistical analysis due to the complex and often unknown relationship between the measurements. Traditional F-test methods for inferring over multiple measures have diminishing power as the number of measures grow, and simply searching over all measures creates a new multiple testing problem. In this work we seek a flexible inferential tool for combination of data. We are motivated by voxel-wise multi-modal group comparisons, e.g. finding where subjects may differ in either BOLD response, as well as by group comparisons over RSNs, e.g. finding differences by searching over the whole brain and a set of some RSNs.

We evaluate the use of a Non-Parametric Combination (NPC) methodology to allow precise inference while implicitly accounting for the dependency structure between modalities, as well as over space.

## 2 The NPC methodology

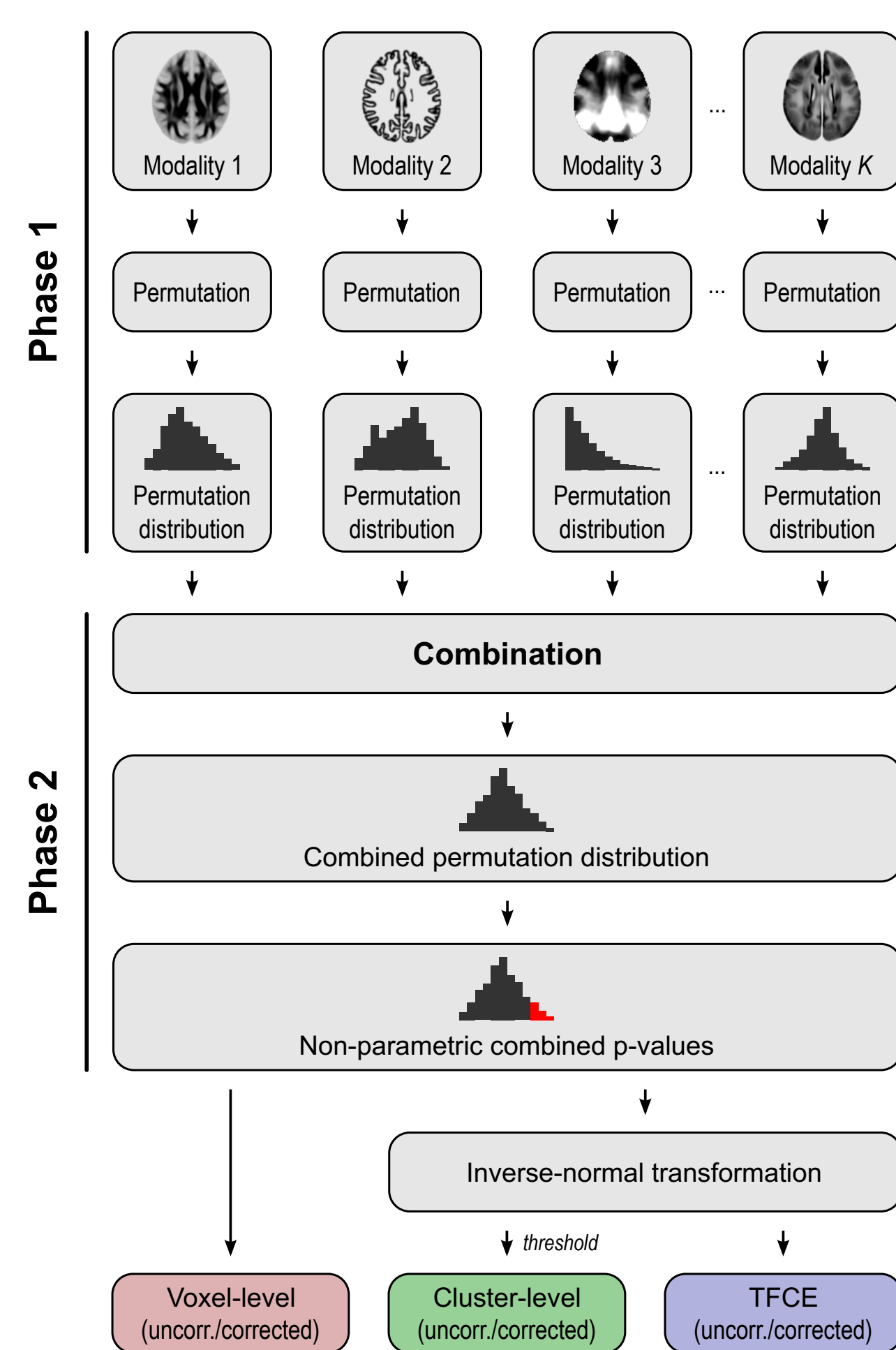
The NPC methodology is a two-phase testing procedure that, under very minimal assumptions, allows multivariate, complex problems to be solved using permutation. The method allows inference on combined data by producing a single measurement that summarises evidence over multiple modalities or RSNs. Relevant features of NPC in the context of neuroimaging include:

- Neither the variables, the model parameters, nor the error terms are assumed to follow any known distribution;
- Dependence structure between the modalities is implicitly accounted for by the method, without the need to be known or modelled.

An overview is shown in **Figure 1**. In the first phase, a *partial test* for each of the methods is performed using synchronized permutations. In the second phase, either the statistics or the p-values for each permutation for all partial tests are combined through a *combination function*. The combination itself produces, for each permutation, a new, combined statistic, which is used to compute the non-parametric p-values for the combined test.

As the result is a p-value (ranging between 0 and 1), a z-score can be produced with the inverse of the normal cumulative distribution function, and conventional cluster-level and TFCE statistics can then also be performed. Voxel-level, cluster-level and TFCE statistics can be corrected to account for multiple testing using the distribution of the largest combined statistic across space.

Figure 1: Overview of the NPC methodology.



### Assumptions

The only assumptions of the NPC procedure is the null-hypothesis exchangeability of the data that goes into each partial test. Exchangeability is a weaker condition than independence, and should be satisfied whenever unrelated subjects are studied.

## Combining functions

Various functions that allow joint inference when multiple hypotheses are being simultaneously tested have been proposed. Most predate the development of the NPC itself. For this study, we evaluated the functions attributed to Tippett, Fisher, Stouffer, Edgington, Friston et al. (these methods test the global null hypothesis) and Nichols et al. (which tests the conjunction null hypothesis).

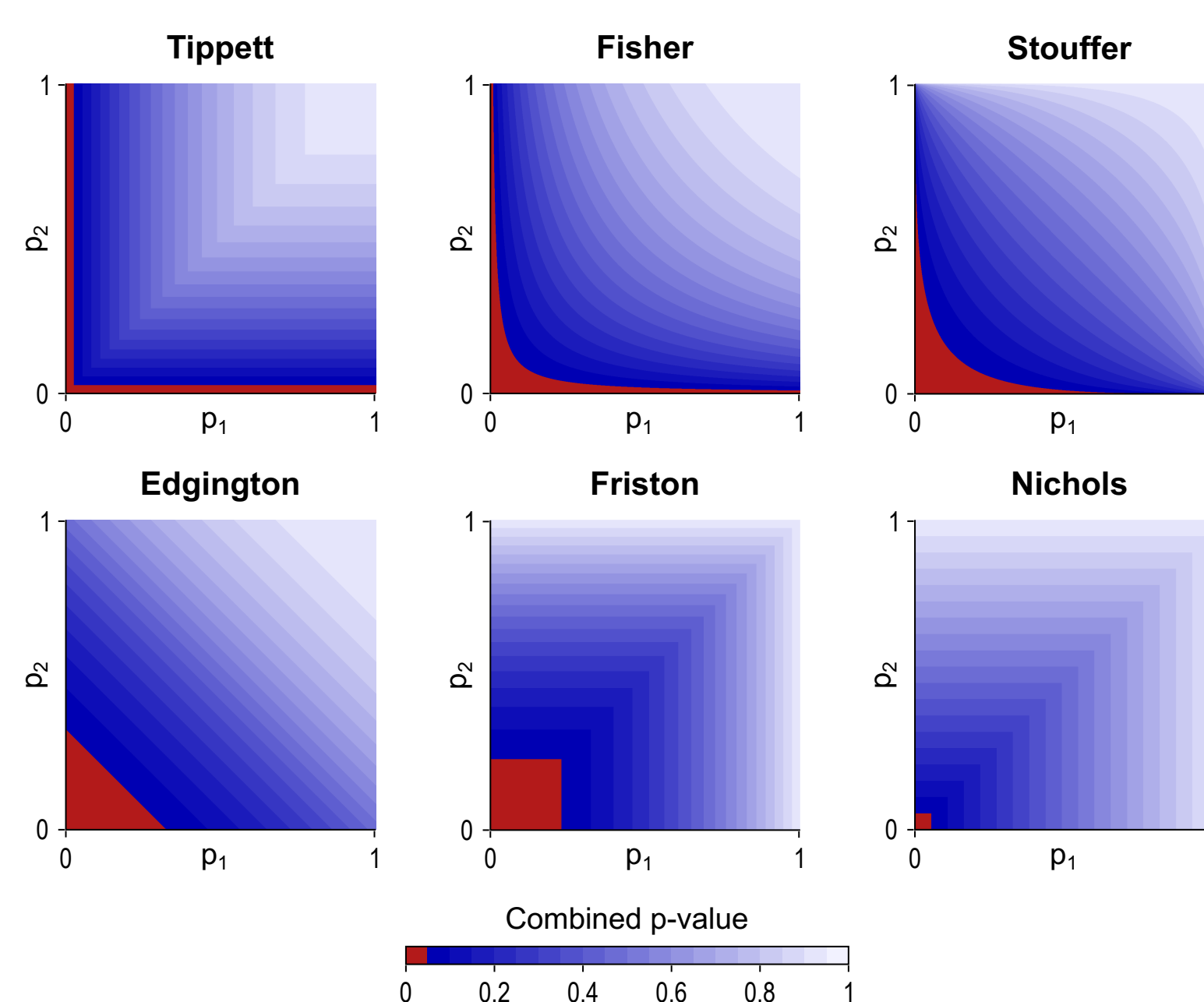
These methods are shown in **Table 1**. Under the assumption of complete independence, significance can be computed analytically from parametric distributions. An illustration of the parametric rejection regions for a simple case of two partial tests is shown in **Figure 2**. Other combining functions can be devised depending on the direction of each of the partial tests [1].

However, if independence between all partial tests is not tenable or, if due to violation of underlying assumptions for the partial tests, their (parametric) p-values are not correct, the analytical p-value obtained for these methods can no longer be considered exact and should be avoided. In these situations, the NPC provides a viable alternative.

Table 1: Combining functions evaluated in this study. The p-values shown are based on independence and standard parametric assumptions. In the NPC, p-values are computed non-parametrically.

Method	Statistic ( $S$ )	Significance (p-value)
Tippett [3]	$\min_k (p_k)$	$1 - (1 - S)^K$
Fisher [4]	$-2 \sum_{k=1}^K \ln(p_k)$	$1 - \chi^2(S; \nu = 2K)$
Stouffer [5]	$\frac{1}{\sqrt{K}} \sum_{k=1}^K \Phi^{-1}(1 - p_k)$	$1 - \Phi(S; \mu = 0, \sigma^2 = 1)$
Edgington [6]	$\sum_{k=1}^K p_k$	$\sum_{j=0}^{\lfloor S \rfloor} (-1)^j \binom{K}{j} \frac{(S-j)^K}{K!}$
Friston et al. [7]	$\max_k (p_k)$	$S^K$
Nichols et al. [8]	$\max_k (p_k)$	$S$

Figure 2: Rejection regions when two tests are combined. The regions in red correspond to a significance level of 0.05.



## Transformation and implementation

As shown in **Figure 1**, the method requires storage of statistical maps for all partial tests, for all permutations. An implementation without such massive storage needs can be considered for methods that allow combination of p-values (as Fisher, Stouffer, Edgington). This consists of using parametric p-values instead of non-parametric for the partial tests at the first phase. This can be seen as a monotonic deterministic transformation of the statistic, in which the p-values lose their original, parametric interpretation, which is present when all the parametric assumptions are satisfied.

Likewise, a transformation can be applied to the combined statistic at the second phase, such that statistic is approximately normal, and so, cluster-level and TFCE can be performed with current tools and well established parameters. This transformation is not necessary for voxel-level inference.

## 3 Evaluation method

Using simulations, we compared the combining functions listed above. The simulations used two groups in a small sample ( $N=14$ , 7 each group) to stress test the methods under low degrees of freedom. 2D images with size  $32 \times 32$  and containing Gaussian or exponential noise, both with zero mean and unit variance, were created. A circular region of signal with peak calibrated at approximately 1.5, and a Gaussian spatial decay with  $\text{FWHM} = 32$  voxels was added at the center of the grid. The number of modalities ( $K$ ) ranged between 1 and 10, with the number of modalities with signal varying between 0 and  $K$  for each. We also evaluated the use of the transformation in the first phase.

## 4 Results

With respect to control over error type I rates, at  $\alpha = 0.05$ , all methods produced exact error rates, indistinguishable from 0.05, as expected given the permutation strategy at the second phase. The only exception was the Nichols method, which is expected to be conservative under the global null tested.

However, we observed that different combining functions show remarkably different sensitivity profiles, which vary according to the number of modalities being combined and with the number of those where true signal is present, as shown in **Figures 3 and 4**. The Tippett and Fisher methods were generally the most powerful, irrespective of the number of modalities being combined, whereas Stouffer and Edgington were only as powerful when signal was present in about half or more of the modalities. The Friston method only had power comparable to the others when signal was present in all modalities, whereas the Nichols method, for being related to a conjunction null test, was the least powerful.

With respect to the transformation of the statistic to a parametric p-value, not only did it provide a fast alternative to the full permutation NPC, but also, for some methods (e.g. Fisher, Stouffer) and in some settings, it caused a slight increase in power.

Figure 3: The power (%) for each method when 10 modalities are combined, as a function of the number of modalities with a signal.

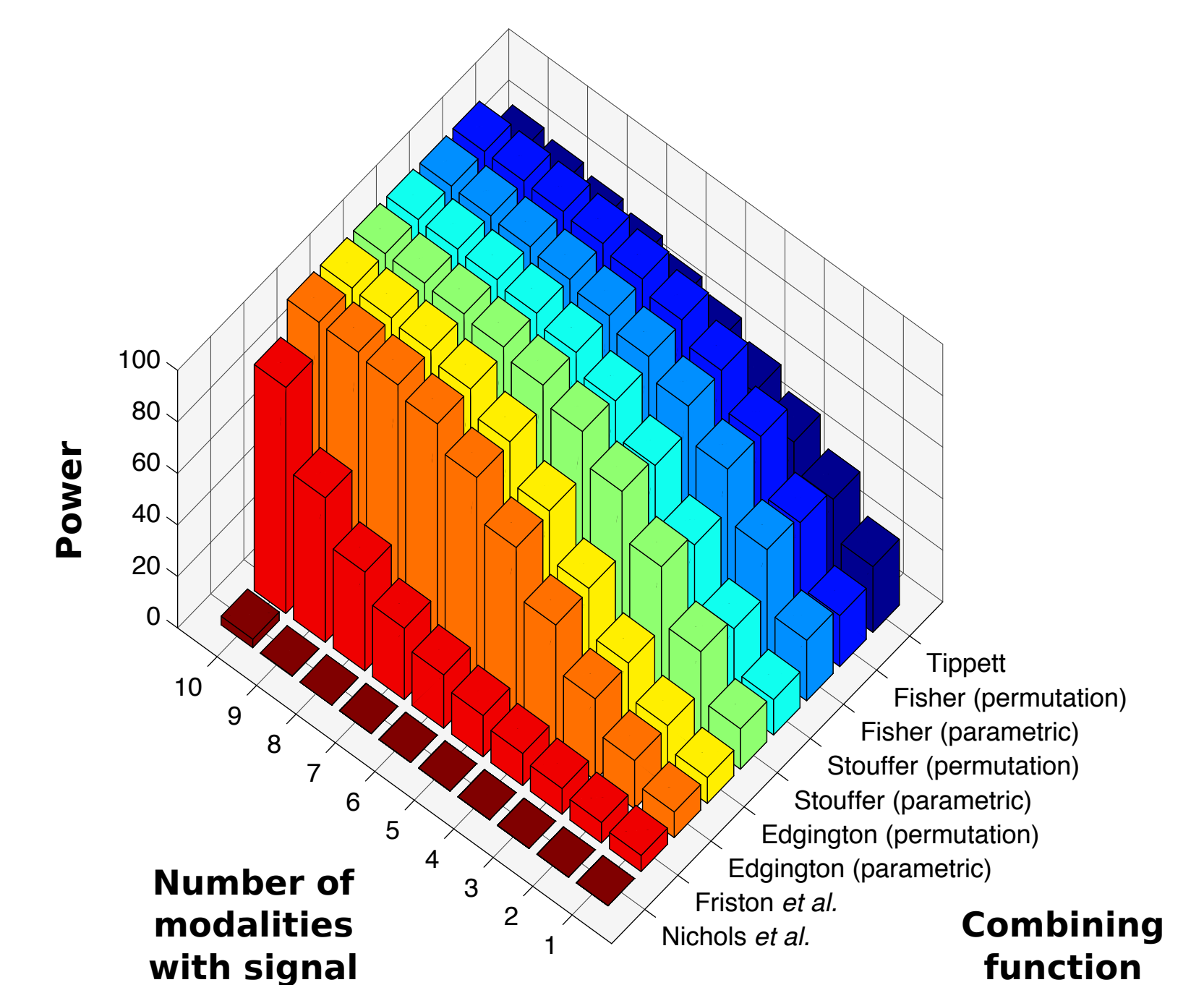
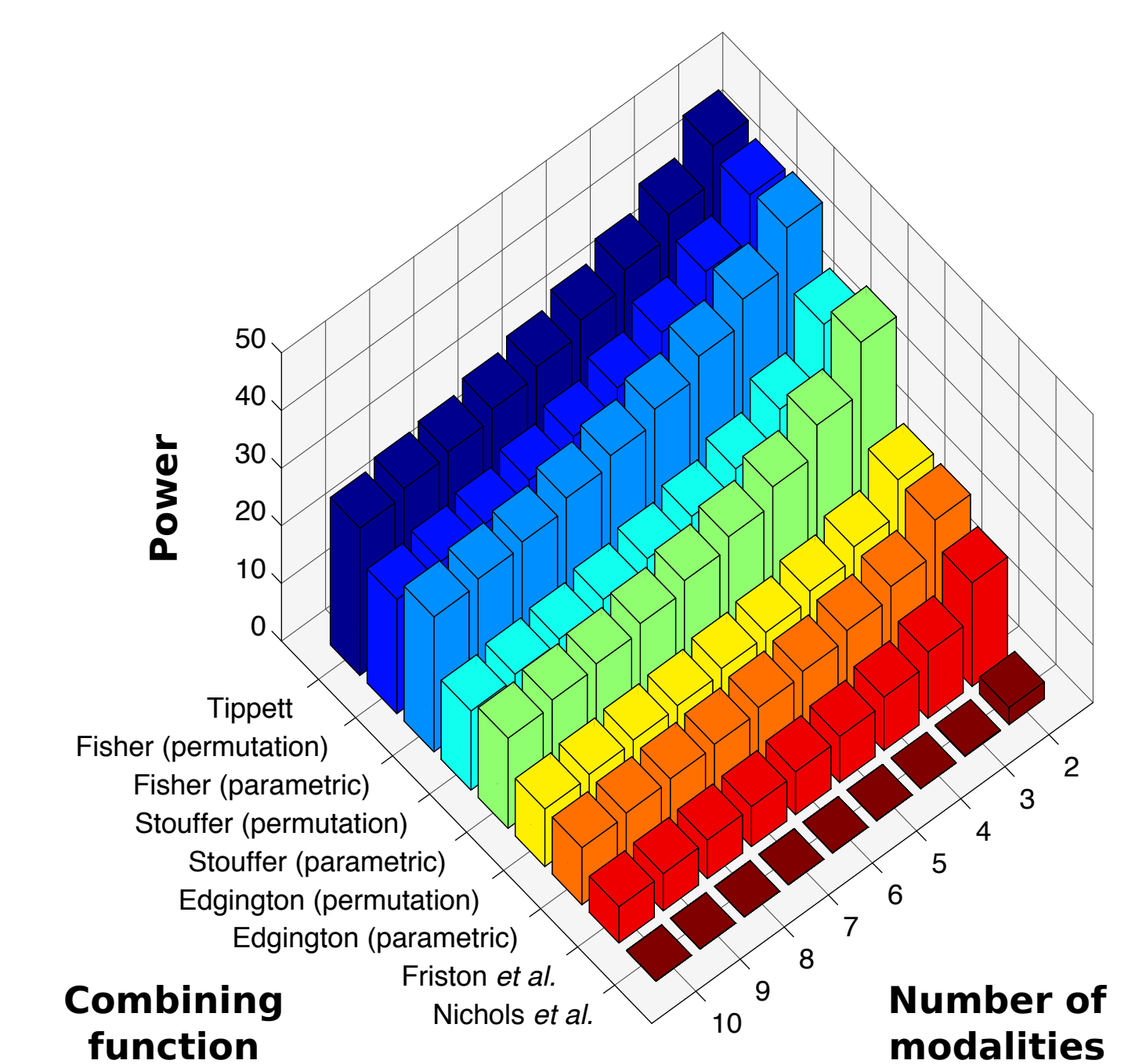


Figure 4: The power (%) for each method when only one modality has signal, as a function of the number of modalities combined.



## 5 Conclusions

- NPC is a viable strategy for analysis of complex designs for imaging applications;
- The choice of the combining function influences power;
- Tippett and Fisher methods are generally the most powerful;
- A fast implementation with parsimonious needs can be attained with the use of the proposed transformation of the statistic;
- The transformation slightly increases power.

## 6 References

- [1] Hayasaka et al. *Neuroimage*. 2006;30(3):768-79; [2] Pesarin & Salmaso. *Permutation Tests for Complex Data*. Wiley, 2010; [3] Tippett. *The Methods of Statistics*. Williams and Northgate, 1931; [4] Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, 1932; [5] Stouffer et al. *Studies in Social Psychology in World War II*. Princeton, 1949; [6] Edgington. *The Journal of Psychology*. 1972;80(2):351-63; [7] Friston et al. *Neuroimage*. 1999;10(4):385-96; [8] Nichols et al. *Neuroimage*. 2005;25(3):653-60.