

This is a signed review by Fred Bookstein.

This manuscript has an excellent point to make, and I would think so even if either

(1) I did not agree with its position, or (2) my own work were not cited so gently and appropriately within it, and even though (3) the title is truly terrible,

since "circularity" has two ENTIRELY different meanings, one having to do with GEOMETRIC circularity and the other with LOGICAL circularity, and the meaning the authors intended is (I trust) the second one whereas the one that instantly leaps to the mind of the typical member of the medical imaging

community must necessarily be the first one, which is to say, the wrong one. The title should instead be something like

Circular reasoning about circles:
logical circularity in voxel-based analysis

or the like.

Anyway, assuming the title is repaired, it is impossible to disagree with the authors' position. My comment is only that this is old news, and that is not meant as a destructive criticism, but only as the suggestion that the literature review be made deeper and wiser than it is.

The paradox here is far from specific to medial imaging. Rather, it is shared by a great many fields that fit complicated, high-dimensional models and then try to draw conclusions from the details of the fitting. The principal idea that drives the solution actually comes from a different field than the authors': this is the idea of the Akaike Information Criterion (AIC) from statistical signal processing theory (see, e.g., Burnham and Anderson's book on model-fitting, second edition, 2002). The Akaike theorem is phrased in terms of likelihood, but the application to least-squares methods is equally felicitous (in linear statistical modelling it is sometimes called the Mallows C-sub-P criterion). The claimed squared error of fit to any multiparameter model has to be inflated by a factor that counts the number of parameters being fitted. In the simplest possible case, a linear regression, the residual sum of squares $N\sigma^2$ of any fit must be increased by precisely 2 --- the integer 2 --- for each additional parameter being fitted before nested models can be compared.

The situation described here in the Tustison manuscript is exactly parallel, and so the classical statistical literature of overfitting may be presumed relevant in exactly its classical form. Suppose the model consists of a circular Gaussian hill of variable center, height, and radius. Presuming a registration on information entirely exogenous to this feature, one would estimate its parameters by some sort of ML computation involving the likelihood of each 3-parameter model in turn at corresponding pixels. In the absence of any such exogenous registration, however, ANY attempt at a local criterion of fit -- the alignment of all of the local peaks of the image, for instance, or the joint centering of apparently circular regions of all

the images -- MUST bias the claimed signal intensity of the resulting analysis in comparison to background. If the task is a signal detection, the data-based registration overstates the confidence of the detection; if the task is the estimate of an amplitude, or the sharpness of a boundary, the registration confound overestimates this.

It is also a near-parallel to the problem identified by Vul in his review of correlation-based methods of partial least squares a couple of years ago. The covariance between singular vectors is necessarily overestimated; the overestimate can be huge, and its unreality correspondingly ludicrous. To understand this problem is to eschew it.

The problem is thus a logical one, not an algorithmic one. If an attempt is to be made to assess the strength of an image signal that is a function of registration, information from that signal MUST be sequestered from the registration procedure. There is just no way around this. Another elementary equivalent to the phenomenon detected by the authors, then, is the familiar context of multiple linear regression. Suppose we have regressed X on Y and are considering whether there is additional information in an additional predictor Z as well. Then, no matter how strong the apparent correlation of the residual of X after regression upon Y, there will be no additional information in Z unless what we are examining is the residual of Z on that same predictor Y. This is, of course, Tukey's method of added-variable plots, a mainstay of any statistician's toolkit. In fact, my own article of 2001, which is appropriately cited by these authors, makes exactly the same point: the information in the registration is confounded with the information in the registered images; you can't say you are testing just one and not the other.

Between Akaike and Tukey, or maybe between Akaike and Bookstein, the literatures of both multivariate statistics and signal processing have been aware for decades of the problem the authors are raising here. That is not an objection to their analysis of the problem nor to their proposed solution (Abstract, line 33: don't do statistical analysis of the same information you used to normalize.) I would suggest only that a revised manuscript acknowledge that the problem on which the authors are focusing is at root a very old one; that it is explicitly the fault of the purveyors of the packages that it continues to arise in academic brain imaging (since, at least symbolically, any system from SPM on could detect the overlap of information between registration calls and peak-detection protocols, and refuse to proceed); and that the only solution is to correct any detected signal for the confounding effects of the registration, in a manner that can easily be done explicitly and algebraically by a variant of the Akaike formula. I happen to think I said most of this already in 2001, but that was before the widespread availability of diffusion tensor data, which looks like it is susceptible to different paradoxes instead. But they are at root the same.

Fred L. Bookstein
University of Vienna,
University of Washington
March 23, 2012