ANDREAS BUJA, DIANE DUFFY, TREVOR HASTIE AND ROBERT TIBSHIRANI

*Bellcore, Bellcore, AT & T Laboratories and University of Toronto*

We would like to congratulate Professor Friedman on this characteristically ingenious advancement in nonparametric multivariate regression modeling. MARS is a triumph of statistical computing and heuristics—the clever algorithmic and heuristic ideas make extensive searching computationally feasible. The resulting modeling technology offers the data analyst a remarkably flexible tool which we found very useful on a difficult real-world problem. We will address a few issues that arose in our reading of this excellent paper and our experience using the MARS program.

**1. Some experience with MARS.** Two of us (Buja and Duffy) acquired some experience with MARS in an extensive analysis of data concerning memory usage in electronic switches. The data comprised 241 observations on 27 variables. It was known from the onset that the available variables gave an incomplete description of the response. Careful and creative regression modeling yielded fits with good global properties ($R^2 \approx 0.995$), but there were still unacceptably large residuals and poor performance on cross-validation tests. The fits which we obtained from MARS, on the other hand, excelled in prediction and cross-validation. In addition, the robustness to influential points which MARS inherits from the local adaptivity of the selected basis functions was very advantageous in our context. Our set of observations was (purposely) chosen to include a subset consisting of notoriously difficult cases. These cases, as expected, wreaked havoc on regression models but MARS was able to adapt to them without degrading the fits to the rest of the cases. In addition, highly accurate MARS models could be built with fewer variables (13 as opposed to 18) which happens to be a true benefit in this situation. The MARS models involved several second and third order interactions which, while impossible to anticipate by subject matter experts, seemed reasonable in the sense that they involved variables which are expected to have large effects on the way memory is used.

An interesting aspect of this analysis is that the data exhibit genuine noise despite the fact that switches are basically deterministic systems. This is because the 27 predictors were selected from a complete set of about 300 predictors based on what information is available to engineers who operate these systems. The success of the MARS fits can only be explained as a result of strong interdependence within the large set of predictors, rendering most of them redundant. Thus, we are profiting from what we call "concurvity" which in most contexts is a cause for concern. Further, models based on the theory of the switching systems would necessarily involve many of the 300 predictors and would therefore be useless to the engineers. One danger in this, or any, data-driven as opposed to theory-driven approach is that the model may be misleading if future predictions occur in areas of the predictor space where data are sparse. It would be useful if MARS were accompanied by diagnostic