Comparing nonparametric surfaces

Adrian W. Bowman
Department of Statistics
The University of Glasgow
Glasgow G12 8QQ
Scotland, U.K.
E-Mail: `adrian@stats.gla.ac.uk`

1. Introduction

Nonparametric smoothing techniques provide well established tools for the description and exploration of trends in data of a wide variety of forms. In recent years, a variety of more formal tests have become available. In the regression setting, additive models, as described by Hastie & Tibshirani (1990), provide a general framework for fitting flexible extensions to standard parametric approaches. The concept of approximate degrees of freedom provides a useful guideline for comparing models although more exact techniques are available in special cases. Simple examples include testing for constant or linear shapes in regression curves and comparing regression curves constructed from different groups of data. Many authors have contributed to this field. Bowman & Azzalini (1997) provide a description of one general approach and references to others.

An important case where inference is useful is in comparing regression surfaces constructed from data in the form of a single response variable y_{ij} and two covariates x_{1ij} and x_{2ij}, where the subscripts i and j denote groups $(i = 1,2)$ and cases $(j = 1,\ldots,n_i)$ respectively. An estimator of each surface at a point of interest (x_1, x_2) is available by solving the local linear regression problem

$$\min_{\alpha,\beta,\gamma} \sum_{j=1}^{n_i} \{y_{ij} - \alpha - \beta(x_{1ij} - x_1) - \gamma(x_{2ij} - x_2)\}^2 \, w(x_{1ij} - x_1; h_1)\, w(x_{2ij} - x_2; h_2).$$

This is a simple extension of the standard local linear approach with one covariate and the calculations can be done very rapidly.

In the simplest case of independent observations one test of equality for this setting follows in a reasonably straightforward manner from well known theory, extending work of Young & Bowman (1995) on nonparametric analysis of covariance. Under the null hypothesis of equality, the biases in the estimation of each surface become equivalent, if appropriate forms of estimation are used and if the smoothing parameters used in each surface are identical. When the surfaces are contrasted these biases therefore cancel. Distributional results are available from well known results on quadratic forms, under the assumption that the original data follow a normal distribution with constant variance.

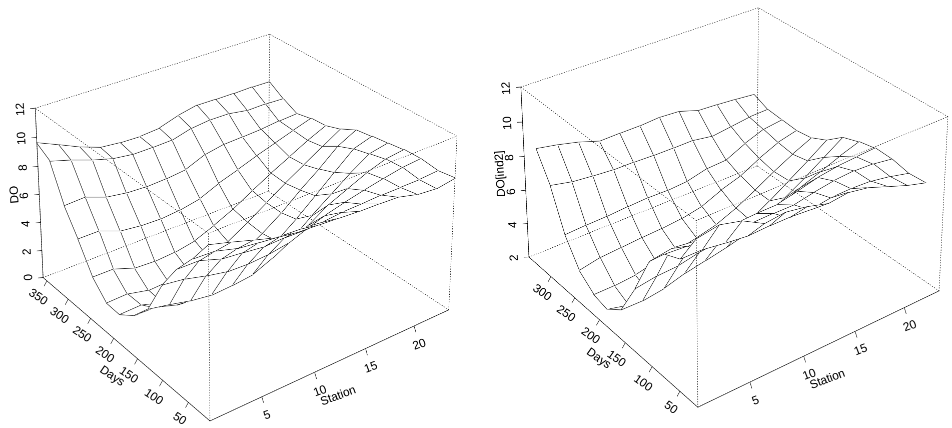

Figure 1: Nonparametric surface plots for the Clyde river data. The left plot refers to years before 1985 while the right plot refers to years after 1985.

2. An example from environmental modelling

In many settings where surface data are recorded, spatial or other forms of correlation may be present. An example is provided by monitoring data for water quality in the River Clyde, collected by the Scottish Environment Protection Agency. Measurements are made of the dissolved oxygen (DO) in the water and interest lies in changes over the years. On a single boat trip several measurements are made at various stations down the river and the DO is expected to change as the river estuary is approached and the proportion of salt water increases. Measurements are also made at different times of year and weather, particularly rainfall, effects are likely to affect the DO values. Figure 1 shows DO surfaces reconstructed from data in the years before and after 1985. A question of interest here is whether there is any indication of differences in these two surfaces, which compare DO after adjusting for station and seasonal effects. There are many detailed aspects which need to be considered in modelling these data, such as the cyclical nature of the seasonal effect. However, for clarity, the surfaces will be discussed here in their simplest form.

The basic form of a global surface comparison is most obviously expressed as

$$\frac{\sum_{i=1}^{2}\sum_{j=1}^{n_i}\{\hat{m}_1(x_{ij}) - \hat{m}_2(x_{ij})\}^2}{\hat{\sigma}^2}.$$

This immediately raises the question of how to estimate the error variance σ^2. Bock *et al.* (2001) discuss this and propose the use of pseudo-residuals from a surface estimator which uses very small smoothing parameters. This provides a two-dimensional analogue of successful estimators constructed in the one covariate case.

The statistic displayed above can be shown to be a ratio of quadratic forms in the data y. The distributional calculations for this statistic then follow those described in detail in other quadratic form problems, as described by Bowman & Azzalini (1997). The complication in this case is that the correlations in the data require a suitable correlation matrix to be incorporated into these calculations. In the present context, it is reasonable to assume that the correlation is present in the repeated observations made down the river on each boat trip. There is no strong evidence of correlation between boat trips which are separated in time. This therefore allows a suitable form of correlation to be adopted and its estimated form used in the distributional calculations.

3. Other forms of surface data

A rather different extension to the standard form of surface comparison arises from stereo-photogrammetry where a pair of cameras and a matching algorithm are used to recover the three-dimensional spatial locations of points on an object. Ayoub *et al.* (1998) describe a system called C3D for taking measurements of this type. The resulting data is of such high resolution that it effectively describes an entire surface, rather than a set of sampled surface points. Collaborative work with the Department of Computer Science and the Dental School of the University of Glasgow is using this system to model the growth of children's faces and to compare the facial surfaces of normal cases with those from cleft lip and palate cases where surgical repair has been carried out.

There already exists a very substantial literature on surface modelling, with particular emphasis on the analysis of an appropriate set of landmarks as well chosen representative points. Dryden & Mardia (1998) give a comprehensive discussion of landmark shape analysis. When the data are collected in surface form, rather than as a set of landmarks, techniques more akin to functional data analysis (Ramsey & Silverman; 1997) become appropriate. The role of principal components and functional data analysis in this setting will be illustrated briefly.

Acknowledgements

The research on environmental surface comparison is being pursued jointly with Marian Scott, Andrew McMullan and Christina Yap and the data were kindly made available by Brian Miller of the Scottish Environment Protection Agency. The work on facial modelling is joint with Mitchum Bock, Shola Ajayi and groups of collaborators in the Department of Computing Science and the Dental School at the University of Glasgow.

REFERENCES

Ayoub, A.F., Siebert, P., Moos, K.F., Wray, D., Urquhart, C. & Niblett, T.B. (1998). A vision-based three-dimensional capture system for maxillo-facial assessment and surgical planning. *British Journal of Oral and Maxillofacial Surgery*, **36**, 353–357.

Bock, M., Beary, I. & Bowman, A.W. (2001). Estimation of error variance in bivariate nonparametric regression. *Technical report, University of Glasgow*.

Bowman, A. & Azzalini, A. (1997). *Applied smoothing techniques for data analysis*. Oxford University Press.

Dryden, I.L. & Mardia, K.V. (1998). *Statistical Shape Analysis*. Wiley: New York.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman & Hall, London.

Ramsay, J.O. & Silverman, B.W. (1997). *Functional data analysis*. Springer-Verlag: New York.

Young, S.G. & Bowman, A.W.(1995). Nonparametric analysis of covariance. *Biometrics*, **51**, 920–931.

RESUME

Il y a beaucoup d'applications importantes où il est intéressant de comparer deux surfaces de régression de forme arbitraire, mais lissées. Il existe des techniques classiques de lissage pour construire de telles surfaces. Un test formel d'égalité peut être construit en étendant les travaux existants sur la comparaison des courbes lissées. Des ajustements appropriés doivent être faits quand les données sont corrélées. Ces idées seront illustrées en utilisant les données rassemblées par l'agence écossaise de protection de l'environnement lors de la surveillance de la qualité d'eau de rivières. Des données sur l'oxygène dissous pour différentes périodes de temps peuvent être comparées en les ajustant en fonction de la période de l'année et de leur emplacement sur la rivière. Une autre forme de comparaison de surfaces, utilisant des données photographiques tridimensionnelles de haute résolution, sera décrite brièvement.