THE MOLLIFIER: A MANAGEMENT TOOL FOR A EUROPEAN SOIL EROSION DATABASE?

by

B.G.J.S. Sonneveld and M.A. Keyzer

Centre for World Food Studies of the Vrije Universiteit (SOW-VU)

Amsterdam, The Netherlands: b.g.j.s.sonneveld@sow.econ.vu.nl

EUROPEAN SOIL EROSION DATA BASE

In the recent past repeated calls were made to organize a European Soil Erosion Database (e.g. Ibañez, 1996; Gabriels, 1998; COST, 1999) so as to ease the incorporation within a common framework of several data sets originating from numerous soil erosion experiments currently conducted in Europe.

Besides offering a full European coverage, such a comprehensive database broadens the empirical basis of erosion models and improves their applicability to local conditions. Furthermore, it facilitates integration with existing spatial and socio-economic data sets.

THE PROBLEM

However, the management of this database, compiled from a wide array of sources is bound to meet with serious difficulties.

First, erosion modellers, who will be the primary suppliers, collect different model variables at varying temporal and spatial scales (Favis-Mortlock, 1998), and their measurements are not evenly spread over the range of variables under observation. Figure 1 illustrates this for an Ethiopian database (SCRP, 2000a-f)[1] by depicting soil loss against LS-factor and a Modified Fournier Index. Second, the functional form of the mathematical relationships of the dependency of erosion on site specific conditions is often unknown.

The software tool described below can assist in addressing these problems. It can detect reliable areas within the data domains, aid in identifying the relationships between variables, and be used for estimation of erosion incidence at unvisited sites or at sites with unreliable data.

[1] The authors would like to thank Professor H. Hurni of the Centre for Development and Environment, University of Berne for allowing the use of the Soil Conservation Research Programme data base.

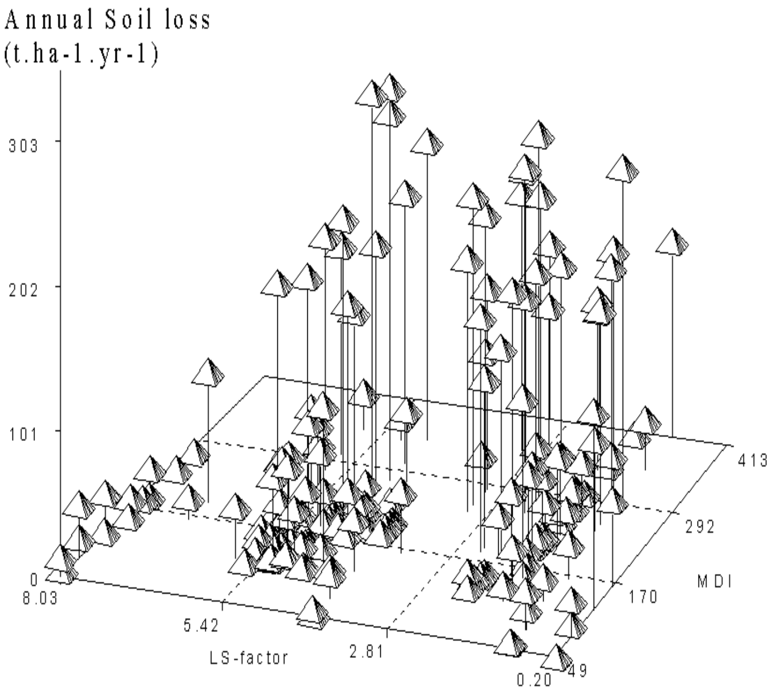


Figure 1 Soil loss against Modified Fournier Index and Slope

NON PARAMETRIC INTERPOLATION

We propose to use the non-parametric interpolation technique of kernel density regression (e.g. Bierens, 1987) that has two major advantages compared to parametric estimation by spline regression or kriging. It fits to the observed data without requiring specification of an explicit functional form. And it generates statistical measures of fit at every point, rather than for the full sample only.

The Mollifier program applies kernel density regression to determine a functional form that can be used for scenario simulation. It offers visual displays of the non-parametric estimates in 3-D graphs while using colour or gray shifts in the regression curve and ground plane to depict associated statistics representing, say, reliable areas in the data domains, quality of fit, and covariates.

MOLLIFIER PROGRAM

The basic principles of calculation may be introduced as follows. Kernel density regression postulates the following stochastic model

$$y(x) = \int y(x+\varepsilon)\,\psi(\varepsilon)\,d\varepsilon\,,$$

where y is the dependent variable (e.g. observed soil loss), x is a vector of explanatory variables and ε denotes measurements errors in x. The function y(x+ε) is the unknown (erosion) function, and the regression takes the expected value of this function. For an infinite sample of observations spread evenly over the domain of x, it would be possible to evaluate this expected value. However, with a finite sample of size S, the value of y can only be estimated and for this the kernel density regression uses the Nadaraya-Watson estimator:

$$\tilde{y}(x) = \sum_s P_s(x)\, y^s,$$

where y^s denote observations and x can be any point. This is a probability weighted sample mean. The probabilities are computed on the basis of the distance of x^s from the given point x, attributing higher weight to nearby points. Weights are assigned using a postulated density function (the kernel) for ε whose spread is controlled by the window size parameter θ. The mollifier program supposes that all the elements of ε are independently and normally distributed. For small samples, a misspecification of this density will affect the estimate but this effect disappears as the sample size becomes larger. The mollifier program implements kernel density regression in a SAS environment. It has a strong graphical orientation and shows the

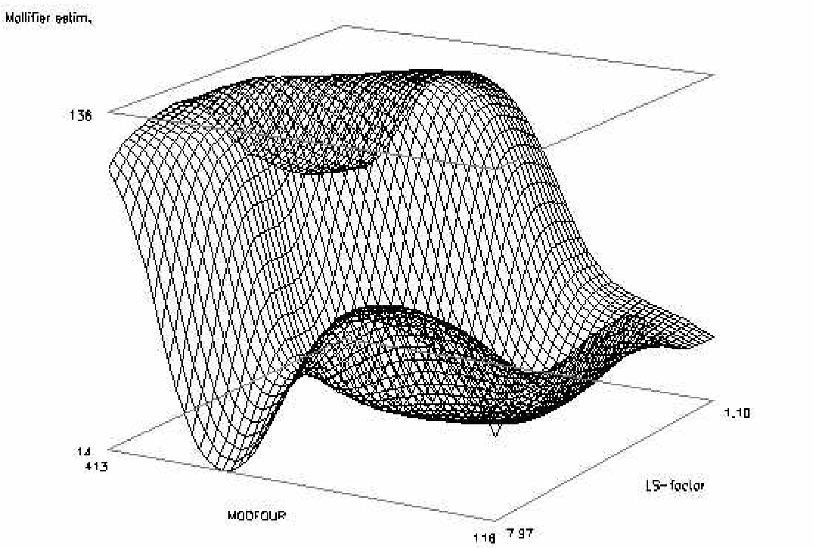


Figure 2 **3-D plot of interpolated soil loss values for the Modified Fournier Index and Slope**

estimated $\tilde{y}(x)$ in 3-D graphs as a surface plot or blanket against two independent variables on, say, a 100×100 grid. Figure 2 gives the results of non-parametric estimates for soil loss for equidistant values along the axis of the LS-factor and MFI. Furthermore, the program can also control for other explanatory variables, keeping them, say, at their sample mean or any other given value. This makes it possible to conduct scenario simulation. In the default mode, the program generates statistics on the reliability of the estimate. A colour shift or shading in the surface plot reflects the likelihood ratio of the observation density, which measures the number of observations on which the function evaluation is based at that point. The colours in a ground plane below the surface plot shows the probability of the actual y falling within a prescribed interval around the mollifier mapping, whose upper and lower bounds are specified as a percentage (default = 10) of the sample mean $\bar{y}$. This eases the task of detecting the influence of outliers and for identifying weak and strong data domains.

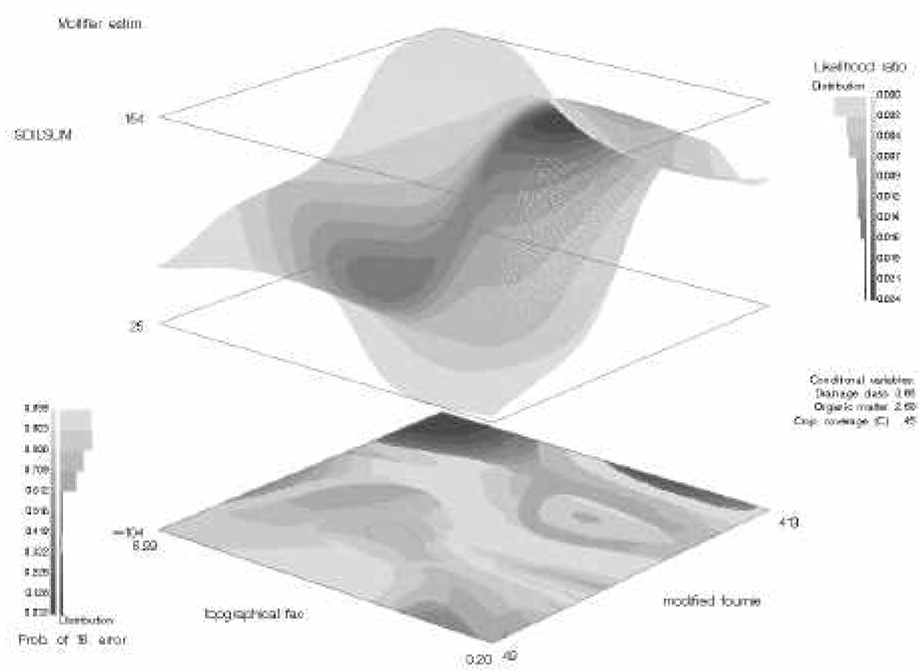
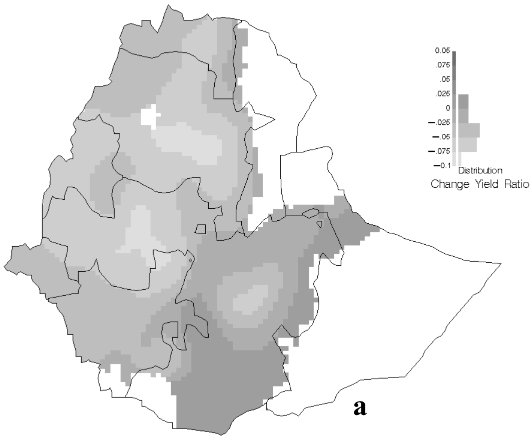
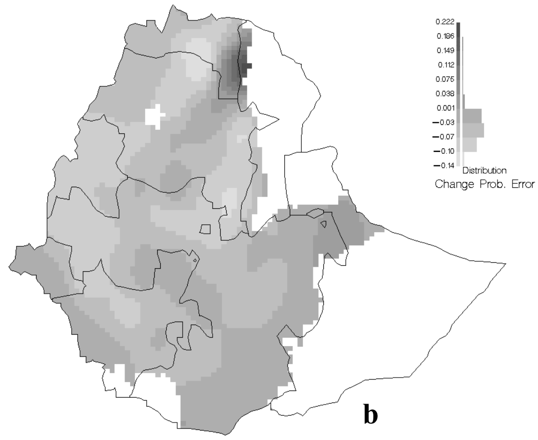


Figure 3 Soil erosion against Modified Fournier index and slopes for annual crops. Covariates: likelihood density and probability of error.

Figure 3 combines these features. It maintains the same variables on the axes as in Figure 2 but the estimates are now being conditioned for mean values of drainage class, organic matter and crop coverage; shades in surface curve and plane show shifts in observation density and probability of error, respectively.

Instead for depicting this statistical information, it is possible to represent co-variates so as to show their location in the dimensions selected. Exchanging the independent variables for longitude and latitude transforms the mollifier picture into a GIS-map while choosing altitude as dependent variable would change the virtual landscapes into a real one. An important distinction from conventional GIS maps is that all spatial values are fully determined by the regression model itself, rather than by some separate interpolation. This permits, for example, to map out directly the outcomes from policy scenarios obtained by fixing exogenous variables at values other than their observed level. Figure 4 shows an application that evaluates the productivity changes in Ethiopia under a scenario of progressive soil degradation and increased





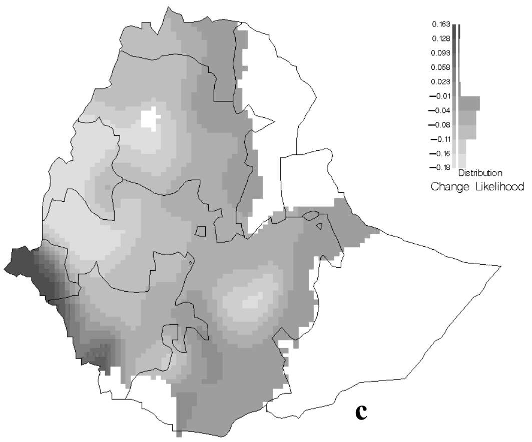


Figure 5. Production changes under progressive land degradation and increased soil fertility in Ethiopia with respect to actual situation (a), likelihood ratio (b) and probability of error (c)

fertilizer use.

Finally, the package can also be used as a flexible curve fitting method to estimate the density function of error terms and estimated coefficients for given parametric forms, without having to impose any normality assumption on their distribution.

APPLICATIONS

The mollifier has proven its usefulness in the characterisation of a wide array of data sets where complex relationships between variables prevail. The method was used in the evaluation of erosion models (Keyzer and Sonneveld, 1998), quantification of land degradation on crop productivity in relation to population density and fertilizer use (Keyzer and Sonneveld, 2000), in yield function analysis (IFAD, 1998), mapping of poverty in Lebanon (IFAD, 2000) and mapping of agricultural production simulations in China (Albersen et al., 2000). It also permits to accommodate qualitative observations based on expert judgements (e.g. Desmet et al., 1995; Gachene, 1995), that may substitute for quantitative data.

CONCLUSION

The mollifier program is a software tool to conduct rapid and interactive exploration of elaborate data sets. It produces both visual displays and tabulated statistical information and could offer a powerful data management tool of a future European Soil Erosion Database.

More information on the mollifier program can be obtained from: Ben Sonneveld, Centre for World Food Studies of the Vrije Universiteit (SOW-VU). De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. b.g.j.s.sonneveld@sow.econ.vu.nl

References

Albersen, P.J., G. Fischer, M.A. Keyzer and L.Sun. 2000. Estimation of agricultural production relations in the LUC model for China. IR-2000-027. IIASA, Austria.

Bierens, H.J. (1987) Kernel density estimations of regression functions. *Advances in Econometrics 6*, Cambridge University Press.

COST (1999) COST 623 home page. Http://www.cost623.leeds.ac.uk/cost623/

Desmet , P.J.J., G.Govers, and D.Goosens (1995) GIS-based simulation of erosion and deposition patterns in: *Experimental geomorphology and landscape ecosystem changes* (J.Poesen and G. Govers eds.).

Gachene, C.K.K. (1995) Evaluation and mapping of soil erosion susceptibility: an example from Kenya. *Soil Use and Management*. 1995, 11: 1, 1-4; 24 ref.

Gabriels D. (1998) Task force on data bank of erosion plot studies in Europe. ESSC newsletter 1+2 pp 25-26.

Favis-Mortlock, D. (1998) The GCTE validation of field–scale soil erosion models. In: Experiences with soil erosion models. (A.Klik eds.). Wiener Mitteilungen. 151. Wien.

Ibañez, J.J. (1996) Proposals for Institutional Consolidation of Demands for Soil Information at a European Level. ESSC Newsletter, 1

IFAD (1998) Rural poverty study of the Caucasus countries. Volumes Armenia, Georgia and Azerbaijan. Rome, Italy

IFAD (2000) Rural poverty assessment and mapping in Lebanon. Rome, Italy

Keyzer, M.A. and P.J. Albersen (1999) Documentation on the Mollifier Program. Centre for World Food Studies of the Vrije Universiteit. Amsterdam, The Netherlands

Keyzer, M.A. and B.G.J.S. Sonneveld (1998) Using the mollifier method to characterize datasets and models: the case of the Universal Soil Loss Equation. *ITC Journal* 1997-3/4.pp 263-272.

Keyzer M.A. and B.G.J.S. Sonneveld (*in press*) The effect of soil degradation on agricultural productivity in Ethiopia: a non-parametric regression analysis. accepted in 'Economic policy reforms and sustainable land use in LDCs,' Springer-Verlag.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*, Chapman and Hall

SCRP (2000) **a)** Andit Tid, Shewa; **b)** Anjeni, Gojam; **c)** Dizi, Illubabor; **d)** Hunde Lafto, Harerge; **e)** Maybar, Wello; **f)** Gununo, Sidamo, Ethiopia: Long-term Monitoring of the Agricultural Environment. Soil Erosion and Conservation Database. Centre for Development and Environment, University of Berne, Switzerland & The Ministry of Agriculture, Ethiopia.

Sonneveld B.G.J.S., M.A. Keyzer and P.J. Albersen. A non-parametric analysis of qualitative and quantitative data for erosion modelling: a case study for Ethiopia. Submitted to proceedings of the International Soil Conservation Organization, Purdue University, Lafayette, USA. 24-28 May 1999. Also as internal Working Paper-99-07.