

Short Communication

Detecting ecological breakpoints: a new tool for piecewise regression

Alessandro Ferrarini

Department of Evolutionary and Functional Biology, University of Parma, Via G. Saragat 4, I-43100 Parma, Italy

E-mail: sgtpm@libero.it, alessandro.ferrarini@unipr.it

Received 7 February 2011; Accepted 13 February 2011; Published online 15 June 2011

IAEES

Abstract

Simple linear regression tries to determine a linear relationship between a given variable X (predictor) and a dependent variable Y. Since most of the environmental problems involve complex relationships, X-Y relationship is often better modeled through a regression where, instead of fitting a single straight line to the data, the algorithm allows the fitting to bend. Piecewise regressions just do it, since they allow emphasize local, instead of global, rules connecting predictor and dependent variables. In this work, a tool called RolReg is proposed as an implementation of Krummel's method to detect breakpoints in regression models. RolReg, which is freely available upon request from the author, could useful to detect proper breakpoints in ecological laws.

Keywords ecological breakpoints; piecewise regression; rolling regression; moving window regression; GIS.

1 Introduction

When modeling the relationship between a single predictor X and a dependent variable Y, a typical regression model takes the form $Y = \alpha + \beta * X$, where Y is expressed as a linear function of X for all pairs of observations (x_i, y_i) .

Contrariwise, the idea behind a piecewise regression is a function that approximates the nonlinearity of a model through the use of separate regression slopes in distinct intervals of the predictor variable domain. In other words, instead of fitting a single straight line to the data, piecewise regression allows the fitting to bend. Krummel et al. (1987) first proposed a piecewise regression which restricts the regression to subsets of the data and performs regression over a window containing a fixed number of data points, with the window ranging over the entire set.

Friedman (1991) proposed an innovative regression algorithm, called MARS, that approximates the nonlinearity of a model by searching over all possible shifts in the relationship between X and Y. MARS has been also used by Ferrarini et al. (2005) to detect proper breakpoints in the area-perimeter relation of landscape patches.

Grossi et al. (2001) suggested a different approach based on an iterative procedure that looks at every possible break of the X-Y relationship to select the best model. This approach presented two drawbacks. When the number of observations is larger than 300, the procedure becomes prohibitive even for the modern computers; in addition, this method requires a large sample size to perform the iterative procedure.

In this work, a tool called RolReg is described. It performs Krummel's method (also known as rolling regression, moving window regression or splitting regression). Although rolling regression is very underused, I suggest that its utilization could be very useful to detect proper breakpoints in ecological laws, such as for instance area-perimeter relation of landscape patches or species-area relationships.

2 Algorithm

Rolling regression performs the following steps:

- a) n pairs of observations are given: $\langle x_i, y_i \rangle$ where $x_i \in X$ and $y_i \in Y$, $1 \leq i \leq n$;
- b) the starting dimension of the moving window (m) and the increment (k) are settled by the user;
- c) rolling regression performs a simple linear regression (SLR) for all pairs of observations (x_i, y_i) where $1 \leq i \leq m$. The m observations within the window are weighted equally;
- d) the slope of regression (β) is plotted against the average value of the x-interval;
- e) the moving window shifts by one observation;
- f) rolling regression performs a SLR for all pairs of observations (x_i, y_i) where $2 \leq i \leq m+1$;
- g) β is plotted against the average value of the x-interval;
- h) the moving window shifts by one observation;
- i) the moving window stops when $i + m \geq n$;
- j) after stopping, a new regression is performed with $m_2 = m + k$ as new window size;
- k) rolling regression performs a SLR for all pairs of observations (x_i, y_i) where $1 \leq i \leq m_2$;
- l) β is plotted against the average value of the x-interval;
- m) the moving window shifts by one observation;
- n) the moving window stops when $i + m_2 \geq n$;
- o) after stopping, a new regression is performed with $m_3 = m_2 + k$ as new window size;
- p) the procedure is repeated g times, where $g = \lceil (n - m) / k \rceil + 1$.

For each window size, the tool RolReg returns a sheet containing the rolling regression graph (Fig. 1). After the passage of g moving windows, the user can look at the g graphs (one for each window size) created by RolReg, and can resolve if one or more breakpoints are proper in the X-Y regression. This step is based on the curve shapes of the g graphs, i.e. an abrupt change found at the same position on graphs generated by different window sizes suggests the existence of a breakpoint. This step is graphical, and does not involve statistical tests.

After the examination of the g curves, the user is asked to enter the number of breakpoints and their values on the x-domain. Given j breakpoints, RolReg fits $j+1$ simple regressions and plots $j+1$ regression lines (Fig. 2).

Last, RolReg generates a final sheet containing the description of the piecewise model (number and location of the breakpoints, equation for each interval) and the comparison with a simple linear regression. The results are fully compatible with any Windows programs.

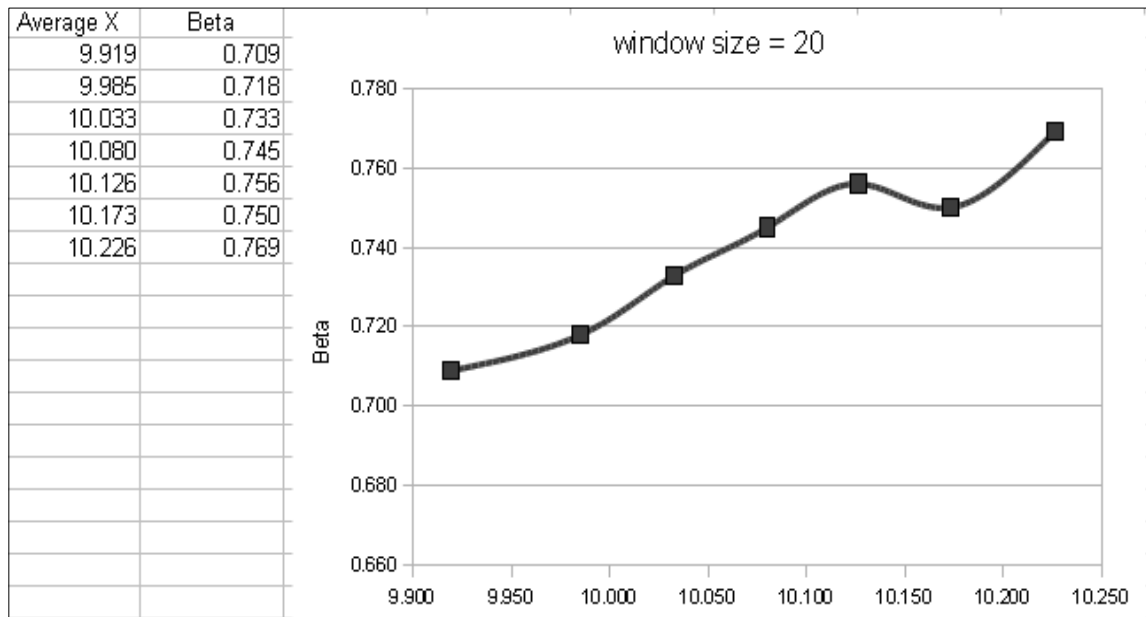


Fig. 1 Results generated by RolReg for a given window dimension.

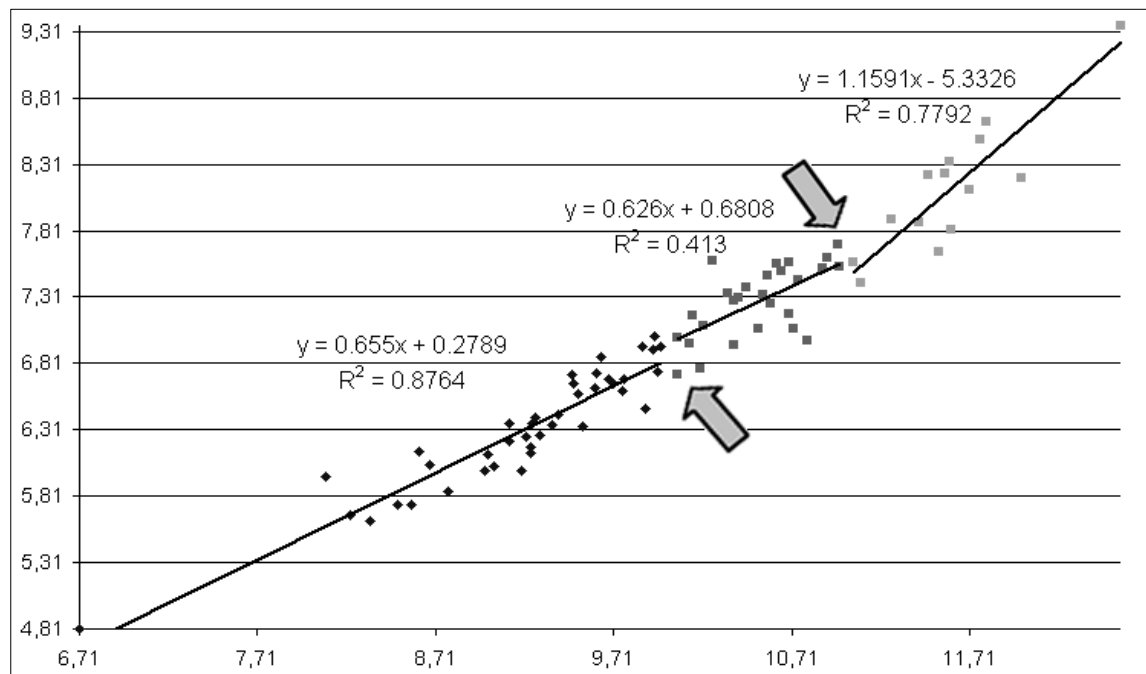


Fig. 2 Rolling regression with 2 breakpoints resulting from the application of RolReg.

3 Conclusions

Environmental data frequently involve situations where one is required to examine the impact of abrupt changes in the data. These changes suggest one or more breakpoints in the law relating predictor and dependent variable. Piecewise regressions are a flexible family of models that can be used to investigate the relationship between two variables. For some researchers this involves situations where the number and location of the breakpoints are known in advance, while others will need to determine the number and location of the breakpoints as part of the estimation process. Rolling regression provides the chance to detect proper breakpoints.

I wrote RolReg as a free add-on for GIS GRASS (Neteler and Mitasova, 2008). It is freely available upon request from the author. As GRASS is very familiar to the most of scientists and students, using RolReg is an easy task as well. I encourage the improvement of both the algorithm and the software.

References

- Ferrarini A, Rossi P, Rossi O. 2005. Ascribing ecological meaning to habitat shape through a piecewise regression approach to fractal domains. *Landscape Ecology*, 20: 799-809
- Friedman JH. 1991. Multivariate adaptive regression splines. *Annals of Statistics*, 19(1): 1-67
- Grossi L, Zurlini G, Rossi O. 2001. Statistical detection of multiscale landscape patterns. *Environmental and Ecological Statistics*, 8: 253-267
- Krummel JR, Gardner RH, Sugihara G, O'Neill RV. 1987. Landscape patterns in a disturbed environment. *Oikos*, 48: 321-324
- Neteler M, Mitasova H. 2008. *Open Source GIS: A GRASS GIS Approach*. Springer, New York, USA