# Accepted Manuscript


Generalized additive partial linear models for analyzing correlated data

Roberto F. Manghi, Francisco José A. Cysneiros, Gilberto A. Paula

PII: S0167-9473(18)30188-9
DOI: https://doi.org/10.1016/j.csda.2018.08.005
Reference: COMSTA 6666

To appear in: *Computational Statistics and Data Analysis*

Received date : 23 March 2018
Revised date : 7 August 2018
Accepted date : 9 August 2018

Please cite this article as: Manghi R.F., Cysneiros F.J.A., Paula G.A., Generalized additive partial linear models for analyzing correlated data. *Computational Statistics and Data Analysis* (2018), https://doi.org/10.1016/j.csda.2018.08.005

# Generalized additive partial linear models for analyzing correlated data

Roberto F. Manghi[b], Francisco José A. Cysneiros[b], Gilberto A. Paula[a,*]

[a] *Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brasil.*
[b] *Departamento de Estatística, Universidade Federal de Pernambuco, Brazil*

**Abstract**

Statistical procedures are proposed in generalized additive partial linear models (GAPLM) for analyzing correlated data. A reweighed iterative process based on the backfitting algorithm is derived for the parameter estimation from a penalized GEE. Discussions on the inferential aspects of GAPLM, particularly on the asymptotic properties of the former estimators as well as on the effective degrees of freedom derivation, are given. Diagnostic methods, such as leverage measures, residual analysis and local influence graphs, under different perturbation schemes, are proposed. A small simulation study is performed to assess the empirical distribution of the parametric and nonparametric estimators as well as of some proposed residuals. Finally, a motivating data set is analyzed by the methodology developed through the paper.

*Keywords:* backfitting algorithm, diagnostic procedures, longitudinal data, natural cubic splines, semiparametric models.

## 1. Introduction

Generalized additive partial linear models (GAPLM) for correlated data (see, for instance, Wang et al., 2014; Lian et al., 2014 and Manghi et al., 2017) comprise an important approach for modeling clustered, repeated measurement and longitudinal data. Such models have the feature of jointly modeling the mean structure by parametric and nonparametric components, with the information only on the marginal distributions as well as on the within-subject correlation structure. The GAPLM class for correlated data combines two well known approaches, generalized estimating equations (GEE) (Liang and Zeger, 1986) and generalized additive models (Hastie and Tibshirani, 1990). Recent reviews on these subjects may be found, respectively, in the books by Hardin and Hilbe (2012) and Wood (2017).

*Corresponding author
*Email addresses:* robertof23@gmail.com (Roberto F. Manghi), cysneiros@de.ufpe.br (Francisco José A. Cysneiros), giapaula@ime.usp.br (Gilberto A. Paula)