Bayesian forecasting of temporal gene expression by using an autoregressive panel data approach

M. Nascimento[1]*, F.F. e Silva[2]*, T. Sáfadi[3], A.C.C. Nascimento[1], L.M.A. Barroso[1], L.S. Glória[2] and B. de S. Carvalho[4]

[1]Departamento de Estatística, Universidade Federal de Viçosa, Viçosa, MG, Brasil
[2]Departamento de Zootecnia, Universidade Federal de Viçosa, Viçosa, MG, Brasil
[3]Departamento de Ciências Exatas, Universidade Federal de Lavras, Lavras, MG, Brasil
[4]Departamento de Genética Médica, Universidade Estadual de Campina, Campinas, SP, Brasil

*These authors contributed equally to this study.
Corresponding author: M. Nascimento
E-mail: moysesnascim@gmail.com

ABSTRACT. We propose and evaluate a novel approach for forecasting gene expression over non-observed times in longitudinal trials under a Bayesian viewpoint. One of the aims is to cluster genes that share similar expression patterns over time and then use this similarity to predict relative expression at time points of interest. Expression values of 106 genes expressed during the cell cycle of *Saccharomyces cerevisiae* were used and genes were partitioned into five distinct clusters of sizes 33, 32, 21, 16, and 4. After removing the last observed time point, the agreements of signals (upregulated or downregulated) considering the predicted expression level were 72.7, 81.3, 76.2, 68.8, and 50.0%,

respectively, for each cluster. The percentage of credibility intervals that contained the true values of gene expression for a future time was ~90%. The methodology performed well, providing a valid forecast of gene expression values by fitting an autoregressive panel data model. This approach is easily implemented with other time-series models and when Poisson and negative binomial probability distributions are assumed for the gene expression data.

Key words: Time series; Temporal gene expression; Posterior predictive distributions