Publications of Manuel Rizzo

This document lists all peer-reviewed publications of Manuel Rizzo, Chair for Algorithm Engineering, Hasso Plattner Institute, Potsdam, Germany. This listing was automatically generated on March 3, 2021. An up-to-date version is available online at hpi.de/friedrich/docs/publist/rizzo.pdf.

Journal articles

[1] Battaglia, F., Cucina, D., Rizzo, M., Parsimonious periodic autoregressive models for time series with evolving trend and seasonality. In: *Statistics and Computing* 30, pp. 77–91, 2020.

This paper proposes an extension of Periodic AutoRegressive (PAR) modelling for time series with evolving features. The large scale of modern datasets, in fact, implies that the time span may subtend several evolving patterns of the underlying series, affecting also seasonality. The proposed model allows several regimes in time and a possibly different PAR process with a trend term in each regime. The means, autocorrelations and residual variances may change both with the regime and the season, resulting in a very large number of parameters. Therefore as a second step we propose a grouping procedure on the PAR parameters, in order to obtain a more parsimonious and concise model. The model selection procedure is a complex combinatorial problem, and it is solved basing on Genetic Algorithms that optimize an information criterion. The model is tested in both simulation studies and real data analysis from different fields, proving to be effective for a wide range of series with evolving features, and competitive with respect to more specific models.

[2] Battaglia, F., Cucina, D., Rizzo, M., Detection and estimation of additive outliers in seasonal time series. In: *Computational Statistics*, pp. 1–17, 2019.

The detection of outliers in a time series is an important issue because their presence may have serious negative effects on the analysis in many different ways. Moreover the presence of a complex seasonal pattern in the series could affect the properties of the usual outlier detection procedures. Therefore modelling the appropriate form of seasonality is a very important step when outliers are present in a seasonal time series. In this paper we present some procedures for detection and estimation of additive outliers when parametric seasonal models, in particular periodic autoregressive, are specified to fit the data. A simulation study is presented to evaluate the benefits and the drawbacks of the proposed procedure on a selection of seasonal time series. An application to three real time series is also examined.

[3] Cucina, D., Rizzo, M., Ursu, E., Multiple changepoint detection for periodic autoregressive models with an application to river flow analysis. In: *Stochastic Environmental Research and Risk Assessment* 33, pp. 1137–1157, 2019.

River flow data are usually subject to several sources of discontinuity and inhomogeneity. An example is seasonality, because climatic oscillations occurring on inter-annual timescale have an obvious impact on the river flow. Further sources of alteration can be caused by changes in reservoir management, instrumentation or even unexpected shifts in climatic conditions. When such changes are ignored the results of a statistical analysis can be strongly misleading, so flexible procedures are needed for building the appropriate models, which may be very complex. This paper develops an automatic procedure to estimate the number and locations of changepoints in Periodic AutoRegressive (PAR) models, which have been extensively used to account for seasonality in hydrology. We aim at filling the literature gap on multiple changepoint detection by allowing several time segments to be detected, inside of which a different PAR structure is specified, with the resulting model being employed to successfully capture the discontinuities of river flow data. The model estimation is performed by optimization of an objective function based on an information criterion using genetic algorithms. The proposed methodology is evaluated by means of simulation studies and it is then employed in the analysis of two river flows: the South Saskatchewan, measured at Saskatoon, Canada, and the Colorado, measured at Lees Ferry, Arizona. For these river flows we build changepoint models, discussing the possible events that caused discontinuity, and evaluate their forecasting accuracy. Comparisons with the literature on river flow analysis and on existing methods for changepoint detection confirm the efficiency of our proposal.

[4] Rizzo, M., Battaglia, F., Statistical and Computational Tradeoff in Genetic Algorithm-Based Estimation. In: *Journal of Statistical Computation and Simulation* 88, pp. 3081–3097, 2018.

When a genetic algorithm (GA) is employed in a statistical problem, the result is affected by both variability due to sampling and the stochastic elements of algorithm. Both of these components should be controlled in order to obtain reliable results. In the present work we analyze parametric estimation problems tackled by GAs, and pursue two objectives: the first one is related to a formal variability analysis of final estimates, showing that it can be easily decomposed in the two sources of variability. In the second one we introduce a framework of GA estimation with fixed computational resources, which is a form of statistical and the computational tradeoff question, crucial in recent problems. In this situation the result should be optimal from both the statistical and computational point of view, considering the two sources of variability and the constraints on resources. Simulation studies will be presented for illustrating the proposed method and the statistical and computational tradeoff question.

[5] Rizzo, M. On Variability Analysis of Evolutionary Algorithm-Based Estimation. In: *Conference of the Classification and Data Analysis Group (CLADAG)*, pp. 237–242, 2017.

A theoretical framework to analyze variability of parametric estimates obtained via Evolutionary Algorithms (EAs) is proposed. The nature of EAs, in fact, introduces a further source of variability, due to stochastic elements of the procedure. A simulation study employing Genetic Algorithms and Differential Evolution is also conducted in oder to make comments on the effect of these stochastic elements on variability.

[6] Rizzo, M., Battaglia, F., On the Choice of a Genetic Algorithm for Estimating GARCH Models. In: *Computational Economics* 48, pp. 473–485, 2016.

The GARCH models have been found difficult to build by classical methods, and several other approaches have been proposed in literature, including metaheuristic and evolutionary ones. In the present paper we employ genetic algorithms to estimate the parameters of GARCH(1,1) models, assuming a fixed computational time (measured in number of fitness function evaluations) that is variously allocated in number of generations, number of algorithm restarts and number of chromosomes in the population, in order to gain some indications about the impact of each of these factors on the estimates. Results from this simulation study show that if the main purpose is to reach a high quality solution with no time restrictions the algorithm should not be restarted and an average population size is recommended, while if the interest is focused on driving rapidly to a satisfactory solution then for moderate population sizes it is convenient to restart the algorithm, even if this means to have a small number of generations.

Conference papers

[7] Friedrich, T., Krejca, M. S., Lagodzinski, J. A. G., Rizzo, M., Zahn, A., Memetic Genetic Algorithms for Time Series Compression by Piecewise Linear Approximation. In: *International Conference on Neural Information Processing (ICONIP)*, pp. 592–604, 2020.

Time series are sequences of data indexed by time. Such data are collected in various domains, often in massive amounts, such that storing them proves challenging. Thus, time series are commonly stored in a compressed format. An important compression approach is piecewise linear approximation (PLA), which only keeps a small set of time points and interpolates the remainder linearly. Picking a subset of time points such that the PLA minimizes the mean squared error to the original time series is a challenging task, naturally lending itself to heuristics. We propose the piecewise linear approximation genetic algorithm (PLA-GA) for compressing time series by PLA. The PLA-GA is a memetic $(\mu + \lambda)$ GA that makes use of two distinct operators tailored to time series compression. First, we add special individuals to the initial population that are derived using established PLA heuristics. Second, we propose a novel local search operator that greedily improves a compressed time series. We compare the PLA-GA empirically with existing evolutionary approaches and with a deterministic PLA algorithm, known as Bellman's algorithm, that is optimal for the restricted setting of sampling. In both cases, the PLA-GA approximates the original time series better and quicker. Further, it drastically outperforms Bellman's algorithm with increasing instance size with respect to run time until finding a solution of equal or better quality – we observe speed-up factors between 7 and 100 for instances of 90,000 to 100,000 data points.

[8] Cucina, D., Rizzo, M., Ursu, E., Identification of Multiregime Periodic Autoregressive Models by Genetic Algorithms. In: *International Conference on Time Series and Forecasting (ITISE)*, pp. 396–407, 2018.

This paper develops a procedure for identifying multiregime Periodic AutoRegressive (PAR) models. In each regime a possibly different PAR model is built, for which changes can be due to the seasonal means, the autocorrelation structure or the variances. Number and lo- cations of changepoints which subdivide the time span are detected by means of Genetic Algorithms (GAs), that optimize an identification cri- terion. The method is evaluated by means of simulation studies, and is then employed to analyze shrimp fishery data.

[9] Battaglia, F., Cucina, D., Rizzo, M., Generalized Periodic Autoregressive Models for Trend and Seasonality Varying Time Series. In: *Scientific Meeting of the Italian Statistical Society (SIS)*, 2018.

Many nonstationary time series exhibit changes in the trend and seasonality structure, that may be modeled by splitting the time axis into different regimes. We propose multi-regime models where, inside each regime, the trend is linear and seasonality is explained by a Periodic Autoregressive model. In addition, for achieving parsimony, we allow season grouping, i.e. seasons may consist of one, two, or more consecutive observations. Identification is obtained by means of a Genetic Algorithm that minimizes an identification criterion.