Research Journal of Applied Sciences, Engineering and Technology 7(21): 4384-4387, 2014
DOI:10.19026/rjaset.7.813
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: June 19, 2012 Accepted: August 28, 2012 Published: June 05, 2014

**Research Article**

# An Overview on R Packages for Seasonal Analysis of Time Series

[1]Haibin Qiu, [2]Ze Chen and [1]Tingdi Zhao
[1]School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China
[2]School of Resource and Safety Engineering, China University of Mining and Technology-Beijing, Beijing 100083, China

**Abstract:** Time series analysis consists of approaches for analysing time series data so thatimportant information and other features can be isolated from the data. Time series forecasting is the use of a model to predict perspective values on the basis of previouly observed values by a model. Statisticians generally use R project or R language, a free and popular programming language and computer software environment for statistical computing and graphics, for developing statistical computer software and data analysis. Plenty of time series display cyclic variation significant as seasonality, periodic variation, or periodic fluctuations in statistics. This study introducesabundant functions in the R packages TSA, marls, depersonalize and season for analyzing seasonal processes of time series, are introduced in this study. Note that R packages marls, depersonalize and season are included in the comprehensive R archive network task view TimeSeries.

**Keywords:** Periodic fluctuations, periodic variation, R project, seasonality, seasonal processes

## INTRODUCTION

In statistics, signal processing, econometrics and mathematical finance, a time series is a sequence of data points, measured characteristically at sequential time instants spaced at regalia time intervals (Box *et al*., 1976; Hamilton, 1994). Methods for analysing time series data are involved in the Time series so as to evacuate valuable statistics and other features from the data (Box *et al*., 1976; Hamilton, 1994). Time series forecasting uses a model to predict the prospect of values accordingto the past observed values. Time series are often drawn by line graphs (Box *et al*., 1976; Hamilton, 1994). R project or R language, which is an useful and free programming language and computer software surroundings for statistical computing and graphics, is widely applied to statisticians for inproving statistical computer software and data analysis (Ripley, 2001; Gentleman *et al*., 2003). Base R ships, which have a lot of functionalities, are helpful for time series, especially in the stats package. Various packages on CRAN with rapidly summarized below follow this.

In statistics, cyclic variation is shown in numerous time series significant as seasonality, periodic variation, or periodic fluctuations. When examing non-seasonal tendencies, seasonal adjustment is used to shedthe seasonal component of a time series. Seasonally adjusted data for unemployment rates always been reported to reveal the basic trends in labor markets (Wallis, 1974; Hillmer and Tiao, 1982). A cycle-stationary process, a signal having statistical properties that vary cyclically with time, can be regarded as multiple interspersed motionless processes (Gardner, 1986; Gardner *et al*., 2006). Seasonal variation is a part of a time series which is described as the regular and predictable movement around the tendency line in no more than one year (Hylleberg, 1992; Barnett and Dobson, 2010). Organizations affected by seasonal variationneed to identify and measure this seasonality to help with planning for temporary increases or decreases in labor requirements, inventory, training, periodic maintenance and so on (Hylleberg, 1992; Barnett and Dobson, 2010). Besidesthese considerations, the organizations should know that if the variations they have experienced have been more or lessexpected given the regularcyclic variations (Hylleberg, 1992; Barnett and Dobson, 2010). The purpose of this study is to review R packages for seasonal analysis of time series.

## R PACKAGES FOR SEASONAL ANALYSIS OF TIME SERIES

Functions in the R packages TSA, marls, depersonalize and season for analyzing seasonal processes of time series, are introduced in this section. Note thatthe comprehensive R archive network task view TimeSeries encloses the R packages marls, depersonalize and season (Hyndman and Zeileis, 2012).

**Corresponding Author:** Tingdi Zhao, School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China

**Seasonal analysis with the R package TSA:** The function acf is modified from the acf function in the stats package (Chan, 2012). The function arima is identical to the arimax function which builds on and expands the capability of the arima function in R stats viapermitting the combination of transfer functions and innovative and additiveoutliers (Chan, 2012). Note in the computation of Akanke's Information Criterion (AIC), the number of parameters excludes the noise variance (Cryer and Chan, 2008; Chan, 2012). This function is deeply builted on the arima function of the stats center of R (Chan, 2012). The R function seasonextract the season information from an equally spaced time series and makes vector of the season information (Chan, 2012).

**Seasonal analysis of health data:** The Rpackage season is for seasonal analysis of health data, containing regression models, time-stratified case-crossover, plotting functions and residual checks (Barnett *et al*., 2012). The R package season aims to fill up a sizable gap in the R software by offering numeroustools for analyzing seasonal data (Barnett *et al*., 2012). The package season includes graphical methods for showing seasonal data and reversion models for checkting and appraisting seasonal patterns (Barnett *et al*., 2012). The regression models can be used to normal, Poisson or binomial dependent data distributions (Barnett *et al*., 2012). Tools can be applied to not only time series data (equally spaced intime) but alsosurvey data (unequally spaced in time) (Barnett *et al*., 2012). Sinusoidal (parametric) seasonal patterns are available incosinor, nscosinor, as well as models for monthly data in monthglm and the case-crossover method to management for seasonality in the R function nonlintestcasecross (Barnett *et al*., 2012). The R function aaftmakes random linear surrogate data of a time series with the same second-order properties (Barnett *et al*., 2012). The AAFT uses phase-scrambling to create a surrogate of the time series, having a similar spectrum and thenparallel second-order statistics (Barnett *et al*., 2012). The AAFT is valuable for checking for non-linearity in a time series and can be used by nonlintest. (Kugiumtzis, 2000; Barnett *et al*., 2012). The R function case cross suits a time-stratified case-crossover to regularly spaced time series data but not suitable for irregularly spaced discrete data (Barnett *et al*., 2012). The case-crossover method compares "case" days when events occurred (e.g., deaths) with control days to look for differences in exposure that might explain differences in the number of cases (Barnett *et al*., 2012). Control days are chosen to be closed to case days, meaning that only recent variations in the independent variables are related (Barnett *et al*., 2012). Any long-term or seasonal variation in the dependent and independent variables can be eliminated by simply comparing recent values, (Barnett *et al*., 2012). The definition of nearby and the seasonal and long-term models in the independent variables determine this elimination (Barnett *et al*., 2012). Control and case days can be contrastedonly when they are in the same stratum (Barnett *et al*., 2012). The strata length, the default value is 28 days, rules the stratum,hence that cases and controls are compared in four week sections (Barnett *et al*., 2012). Smaller stratum lengths provide a closer control for season, but then decrease the functional number of controls (Barnett *et al*., 2012). Control days,closing to the case day, may have comparable levels of the independent variables (Barnett *et al*., 2012). It is likelytoput an exclusion around the casesto reduce this correlation (Barnett *et al*., 2012). It is possible to additionally match by day of the week (matchdow) to delete any confounding by day of the week, regardless of the fact that this often reduces the number of available controls (Barnett *et al*., 2012). This matching is in exclusive of the strata matching (Barnett *et al*., 2012). To eliminate its effect, it is probable to additionally match case and control days by an influential on-founder (matchconf) (Barnett *et al*., 2012). The method utilizes conditional logistic regression (see coxph and so the parameter estimates are odds ratios (Barnett *et al*., 2012). The code presumes that the data frame includes a date variable (in Date format) called date (Janes *et al*., 2005; Barnett and Dobson, 2010; Barnett *et al*., 2012). The function ciPhase computes the mean and confidence interval for the phase based on aseries of MCMC samples (Barnett *et al*., 2012). The evaluation of the phase are rotated to have a centre of n, the point on the circumference of a unitradius circle that is furthest from zero (Barnett *et al*., 2012). The mean and confidence interval are computed on the rotated values and then the estimates are rotated backward. Fisher (1996), Barnett and Dobson (2010) and Barnett *et al*. (2012). The function nonlintest realizes a bootstrap test of non-linearity in a time series, which uses the third-order moment (Barnett *et al*., 2012). The aaft is used to make linear surrogates with the same second-order properties, but no (third-order) non-linearity (Barnett *et al*., 2012). The third-order moments (third) of these linear surrogates and the actual series are then comparedfrom lags 0 up to *n.lag* (excluding the skew at the co-ordinates (0, 0)) (Barnett *et al*., 2012). The bootstrap test applies on the whole area outside the limits and provides an indication of the overall non-linearity (Barnett *et al*., 2012). The plot using region reveals those co-ordinates of the third order moment that beyonded the null hypothesis limits and can be a valuable clue for guessing the type of non-linearity (Barnett *et al*., 2012). A non-stationary seasonal pattern changes over time; so this model offers potentially extremly elastic seasonal assessments (Barnett *et al*., 2012). The cycle controlles thefrequency of the seasonal estimates and ought to be specified in units of time (Barnett *et al*., 2012). The estimates are made using a forward and backward sweep of the Kalman filter (Barnett *et al*., 2012). Repeated estimates are made using Markov Chain Monte Carlo (MCMC) (Barnett *et al*., 2012). For this reason the model can take a long time to run (we will perfect this in the next version) (Barnett *et al*., 2012). A reasonably long-termmodel should be applied (niters) and the probably inadequate initial estimates should be abandoned (burn in) to give stable estimates. (Barnett and Dobson, 2004, 2010; Barnett *et al*., 2012). The function phasecalc computes the phase given the estimated sine and cosine values from a cosinor model (Barnett *et al*., 2012) and returns the phase in radians, in the range (0, 2) (Barnett *et al*., 2012). The phase is at the top of the sinusoid (Fisher, 1996; Barnett *et al*., 2012). The function plot Circular applies a circular plot helpful for visualising monthly or weekly data (Barnett *et al*., 2012). The length of the variable area 1 is determined by the number of segments (Barnett *et al*., 2012). The plots are also called rose diagrams, with the segments then called 'petals' (Fisher, 1996; Barnett *et al*., 2012). The function plot Month plots results by month (Barnett and Dobson, 2010; Barnett *et al*., 2012). For modelling seasonal data, sinusoidal curves are highly prized (Barnett *et al*., 2012). The function sinusoid plots a sinusoid from 0 to 2 (Barnett *et al*., 2012). West for a seasonal pattern in Binomial data, a test of whether monthly data has a sinusoidal seasonal pattern, has little influence compared with the cosinor test. Walter and Elwood (1975), Barnett and Dobson (2010) and Barnett *et al*. (2012). The function nscosinor will be made faster and the plots are to be improved in the next version, (Barnett *et al*., 2012).

**Multiplicative AR (1) with seasonal processes:** Multiplicative AR (1) (Yamamoto, 1982) with Seasonal processes, processes (Paramonov, 2012), is a stochastic process model built on top of AR (1) and can be abbreviated as MAR(1)S (Petruccelli and Woolford, 1984; Chan *et al*., 1985; Chan and Wei, 1987; Baltagi and Wu, 1999). The R package mar1s (Paramonov, 2012) picks up MAR(1)S or GARCH models (Drost and Nijman, 1993; Lamoureux and Lastrapes, 1990; Nelson and Cao, 1992; Karolyi, 1995; Baillie *et al*., 1996; Tse, 2000; Garcia *et al*., 2005). The package mar1s offers the following procedures for MAR(1)S:

- Fitting
- Composition
- Decomposition
- Advanced simulation
- Prediction (Paramonov, 2012)

The function fit.mar1s fits MAR(1)S process model to time series (Paramonov, 2012). The function composes. Ar1 composes AR (1) process realization by given vectors of innovations (Paramonov, 2012). The function compose.mar1s composes MAR(1)S process realization by given vector of log-innovations (Paramonov, 2012). The function decomposes. Ar1 extracts AR (1) process residuals from time series (Paramonov, 2012). The function decompose. mar1s extracts MAR(1)S process components from time series (Paramonov, 2012). The function sim.mar1s simulates from MAR(1)S process (Paramonov, 2012). The function predict.mar1s is a wrapper around the functionsim.mar1s which estimates confidence intervals for the future values of the MAR(1)S process (Paramonov, 2012). The function seasonal.ave extracts seasonal component of time series by averaging observations on the same place in the cycle (Paramonov, 2012). The function seasonal. smooth extracts seasonal component of time series by fitting the data with a linear combination of smooth periodic functions (Paramonov, 2012).

**Optimal depersonalization for geophysical time series using AR fitting:** A harmonic regression is fit to the data to estimate the seasonal means and standard deviations (Young *et al*., 1999; McLeod, 2012). The number of terms in the harmonic regression may be determined using the Bayesian Information Criterion (BIC) or generalized Akaike's Information Criterion (generalized AIC) (Young *et al*., 1999; McLeod, 2012; Sakai, 1990; Haughton *et al*., 1990; Albert and Hansberger, 2005; Tominaga, 2010). The R package deseasonalize that is an optimal depersonalization for geophysical time series using AR fitting, depersonalize daily or monthly time series (McLeod, 2012; Hipel and McLeod, 1994; McLeod and Zhang, 2008). In R package depersonalizethe main function ds is a depersonalization process for monthly and annual (McLeod, 2012; Hipel and McLeod, 1994; McLeod and Zhang, 2008). Most users would employ the ds, but not the utility function getds in R package depersonalize (McLeod, 2012). The function print.depersonalizecan give a terse summary (McLeod, 2012). The function summary. Depersonalize provides summary for depersonalize objects (McLeod, 2012; Hipel and McLeod, 1994; McLeod and Zhang, 2008).

## CONCLUSION

Time series analysis includes methods for analysing time series data so as to deracinate meaningful statistics and other characteristics from the data. Plenty of functions in the R packages TSA, mar1s, depersonalize and season for analyzing seasonal processes of time series, are introduced in this study.

## REFERENCES

Albert, P. and S. Hunsberger, 2005. On analyzing circadian rhythms data using nonlinear mixed models with harmonic terms. Biometrics, 61: 1115-1120.

Baillie, R., C. Chung and M. Tieslau, 1996. Analysing inflation by the fractionally integrated arfima-garch model. J. Appl. Econ., 11: 23-40.

Baltagi, B. and P. Wu, 1999. Unequally spaced panel data regressions with AR(1) disturbances. Economet. Theor., 15: 814-823.

Barnett, A. and A. Dobson, 2004. Estimating trends and seasonality in coronary heart disease. Stat. Med., 23: 3505-3523.

Barnett, A. and A. Dobson, 2010. Analysing Seasonal Health Data. Springer Verlag, ISBN: 978-3-642-10748-1.

Barnett, A., P. Baker and M. Barnett, 2012. Package 'season'. Retrieved from: http://cran.r-project.org/web/packages/season/season.pdf.

Box, G., G. Jenkins and G. Reinsel, 1976. Time Series Analysis: Forecasting and Control. Revised Edn., Holden-Day, Oakland, California.

Chan, K.S., 2012. Package 'TSA'. Retrieved from: http://cran.rproject.org/ web/packages/TSA/TSA.pdf.

Chan, N.H. and C.Z. Wei, 1987. Asymptotic inference for nearly nonstationary AR(1) processes. Ann. Stat. 15(3): 1050-1063.

Chan, K.S., J.D. Petruccelli, H. Tong and S.W. Woolford, 1985. A multiple threshold AR(1) model. J. Appl. Probab., 22(2): 267-279.

Cryer, J. and K. Chan, 2008. Time Series Analysis: With Applications in R. 2nd Edn., Springer, New York, pp: 491, ISBN: 0387759581.

Drost, F. and T. Nijman, 1993. Temporal aggregation of garch processes. Econ. Soc. J. Econ., 61(4): 909-927.

Fisher, N., 1996. Statistical Analysis of Circular Data. Cambridge University Press, Cambridge, pp: 296, ISBN: 0521568900.

Garcia, R., J. Contreras, M. Van Akkeren and J. Garcia, 2005. A garch forecasting model to predict day-ahead electricity prices. IEEE T. Power Syst., 20(2): 867-874.

Gardner, W., 1986. The spectral correlation theory of cyclostationary time-series. Signal Process., 11(1): 13-36.

Gardner, W., A. Napolitano and L. Paura, 2006. Cyclostationarity: Half a century of research. Signal Process., 86(4): 639-697.

Gentleman, R., R. Ihaka *et al*., 2003. The R Project for Statistical Computing. Retrieved from: http://www.r-project.org/.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton, pp: 820, ISBN: 0691042896.

Haughton, D., J. Haughton and A. Izenman, 1990. Information criteria and harmonic models in time series analysis. J. Comput. Simulat., 35: 187-207.

Hillmer, S. and G. Tiao, 1982. An arima-model-based approach to seasonal adjustment. J. Am. Stat. Assoc., 77(377): 63-70.

Hipel, K.W. and A.I. McLeod, 1994. Time Series Modelling of Water Resources and Environmental Systems. Elsevier, Amsterdam, New York, pp: 1013, ISBN: 0444892702.

Hylleberg, S., 1992. Modelling Seasonality. Oxford University Press, New York, pp: 476, ISBN: 019877317X.

Hyndman, R.J. and A. Zeileis, 2012. Contact: Rob.Hyndman@monash.edu Version: 2012-04-25.

Janes, H., L. Sheppard and T. Lumley, 2005. Case-crossover analyses of air pollution exposure data: Referent selection strategiesand their implications for bias. Epidemiology, 16(6): 717-726.

Karolyi, G., 1995. A multivariate garch model of international transmissions of stock returns and volatility: The case of the United States and Canada. J. Bus. Econ. Stat., 13(1): 11-25.

Kugiumtzis, D., 2000. Surrogate data test for nonlinearity including nonmonotonic transforms. Phys. Rev. E, 62(1): 25-28.

Lamoureux, C. and W. Lastrapes, 1990. Persistence in variance, structural change and the garch model. J. Bus. Econ. Stat., 8(2): 225-234.

McLeod, A. and Y. Zhang, 2008. Improved subset autoregression: With R package. J. Stat. Softw., 28: 1-28.

McLeod, A.I., 2012. Package 'deseasonalize'. Retrieved from: http://cran.r-project.org/web/packages/deseasonalize/deseasonalize.pdf.

Nelson, D. and C. Cao, 1992. Inequality constraints in the univariategarch model. J. Bus. Econ. Stat., 10(2): 229-235.

Paramonov, A., 2012. Package 'mar1s'. Retrieved from: http://cran.r-project.org/web/packages/mar1s/mar1s.pdf.

Petruccelli, J. and S. Woolford, 1984. A threshold AR(1) model. J. Appl. Probab., 21(2): 270-286.

Ripley, B., 2001. The R project in statistical computing. MSOR Connections. Newsletter LTSN Maths. Stats Network, 1: 23-25.

Sakai, H., 1990. An application of a BIC-type method to harmonic analysis and a new criterion for order determination of an AR process. IEEE T. Acoust. Speech, 38: 999-1004.

Tominaga, D., 2010. Periodicity detection method for small-sample time series datasets. Bioinform. Biol. Insigh., 4: 127-136.

Tse, Y., 2000. A test for constant correlations in a multivariate garch model. J. Econometrics, 98: 107-127.

Wallis, K., 1974. Seasonal adjustment and relations between variables. J. Am. Stat. Assoc., 69(345): 18-31.

Walter, S. and J. Elwood, 1975. A test for seasonality of events with a variable population at risk. Brit. J. Prev. Soc. Med., 29: 18-21.

Yamamoto, K., 1982. Yearly and seasonal process of maturation of ray parenchyma cells in pinus species. Res. Bull. Extl. For. Hokkaido Univ., 39: 245-296.

Young, P., D. Pedregal and W. Tych, 1999. Dynamic harmonic regression. J. Forecasting, 18: 369-394.